AIR FORCE INSTITUTE OF TECHNOLOGY

AIR UNIVERSITY
UNITED STATES AIR FORCE

SCHOOL OF ENGINEERING

WRIGHT-PATTERSON AIR FORCE BASE, OHIO

AN ANALYSIS OF FIRST UNIT
LABOR COSTS FOR FIXED
WING AIRCRAFT

THESIS

GSA/SM/70-17 Roger M. Smith
Major USAF

GSA/SM/70-17

AN ANALYSIS OF FIRST UNIT LABOR COSTS FOR FIXED WING AIRCRAFT

THESIS

Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

Air University

in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

by

Roger M. Smith, B.A.
Major USAF

Graduate Systems Analysis

June 1970

This document is subject to special export controls and each transmittal to foreign governments or foreign nationals may be made only with prior approval of the Dean of Engineering, Air Force Institute of Technology (AFIT-SE), Wright-Patterson AFB, Ohio 45433

Preface

When this study was first formulated, it was hoped that it could examine aircraft of several types built by several manufacturers. It was felt that the results of a broad study would have extensive application within the air force and the airframe industry. It was a disappointment to find only one manufacturer willing to provide information necessary to support this study. Even though he requires anonymity, without his information this study would have been impossible, and I wish to express my thanks for his cooperation and my hopes that he will find the ideas contained in this study useful.

I am also deeply grateful to the faculty of the Systems Management Department of the School of Engineering in the Air Force Institute of Technology for allowing me to develop this study in my own way. Though it is not as learned as it would be if each step were directed by a professor, the results appear useful, and I have benefited greatly from the nearly ideal learning situation they have created. More direct involvement on the faculty's part could have reduced the number of dead ends pursued and tangents explored, but this would have been at the expense of my profiting from my own mistakes.

Roger M. Smith

Contents

List of Figures

List of Tables

Definitions

AMPR Weight. The empty weight of the aircraft less engines, wheels, tires, fuel cells, electronic equipment, instruments, and other equipment as defined in the Aeronautical Manufacturers' Planning Report (Ref 28:4-6).

Critical Point. If the labor costs for each aircraft are plotted against some performance or engineering variable, one point will appear more discontinuous than the others. This discontinuity might be small or large. The aircraft representing the point will vary depending upon the specific variable against which labor costs are being compared. For any particular variable, the most discontinuous point is designated the critical point.

Labor Costs. Labor costs will be used as a general term to refer to a quantity of labor resource that must be expended to perform a task. Whether the quantity is measured in hours, hours per pound, or dollars is unimportant. All of these units of measure, and others, can be used to describe the same labor expenditure.

Labor Hours. The number of direct labor manhours required to manufacture an airframe and to install equipment necessary to transform the airframe into a flyable aircraft. This study uses without modification the detailed definition of direct labor hours used in the accounting system of the manufacturer who supplied the data for this study. This definition conforms generally to the definition of direct labor hours given in Asher (Ref 3:48-50).

Maximum Gross Weight. An aircraft's maximum allowable gross weight for take-off for normal operation under standard atmospheric conditions.

Maximum Speed. The maximum true airspeed in knots attainable at any altitude regardless of whether the limiting factor is lack of additional power or structural limitations.

Symbols

H A general symbol for the total number of hours required to manufacture the first unit. H is found by multiplying a specific Y by its corresponding AMPR weight.

H_1 Y_1 multiplied by AMPR weight.

H_a Y_a multiplied by AMPR weight.

r Learning fraction. A number between zero and one which gives the rate of reduction in labor requirements that occur in each ten year period.

R^2 Coefficient of determination.

R_i^2 The partial coefficient of determination for the ith independent variable in the equation being examined.

T A general symbol for time.

T^{exp} Exponential time. A time scale found by subtracting some base year from the start production dates of all the aircraft in the study. The name exponential is used to distinguish this time scale and its use from straight line time below. When this time scale is used in log regressions, the resulting regression coefficients are exponents after the inverse logarithmic transformation, hence the name. When a specific value of the exponent is being considered, the term "exp" will be replaced by a number, i.e., $T^{-.3}$ is the time

scale with each of its terms raised to the -.3 power.

T_r Straight line time. A time adjustment factor that will reduce a quantity to a certain fraction of its original value for each ten years being considered. When a specific learning fraction is being considered, the learning fraction "r" will be replaced with a number, i.e., $T_{.8}$. See chapter V, page 49.

$T_{.8}$ Straight line time adjustment factor using r = .8 and base year equal to the start production date of the last aircraft in the sample. See chapter V, page 50.

V Maximum speed.

W_a AMPR weight.

W_g Maximum gross weight.

W_{g-a} Maximum gross weight minus AMPR weight.

Y The symbol Y will be used as a general symbol to represent a number in direct labor hours per AMPR pound of aircraft weight. The number will always be obtained from the regression equation that best fits the program in question. When subscripted, it will represent the number of hours per pound to produce the cumulative unit designated by the subscript. (i.e., Y_{100} refers to the value given by the regression line to produce the 100th unit)

Y_1 The parameter of the learning curve equation found by regressing the average labor hours per pound per aircraft in all lots produced before unit 100 on the cumulative lot mid points. When the program in question had a prototype, the prototype lot was excluded and counting began with the first lot after the prototype.

Y_a The parameter of the learning curve equation found in the same manner as Y_1 except an adjustment factor of from 0.0 to 1.0 is added to all cumulative lot mid points before regression. See chapter III.

Y_{a-1} The value of Y_a found by extrapolating from lot 1 mid point to the Y axis using the slope of the Y_a learning curve equation.

Y* A value in hours per AMPR pound for the first unit labor costs found by inserting the independent variables for the aircraft in question in an estimating equation and solving for Y. Y* is an equation produced estimate of either Y_1 or Y_a.

Y*-Y The deviation between equation produced estimate of Y and the actual value of Y used to form the estimating equation.

Abstract

This study examines the first unit labor cost parameter of the learning curve equations belonging to seven different aircraft. The purpose of this investigation is to find a method which will produce good estimates of this parameter prior to the start of production of an aircraft.

It was found that simple linear, and log linear, multiple variable relations could not provide an accurate estimate. However, non-linear functions of weight and time were able to estimate historical data within 4% of the actual value. It is concluded that equations of this form should lead to very accurate estimates of labor cost.

AN ANALYSIS OF FIRST UNIT LABOR COSTS FOR FIXED WING AIRCRAFT

I. Introduction

This analysis concerns estimating the total labor costs to manufacture the first airframe of a particular aircraft type; i.e., the total labor costs for the first B-52A produced. It does not concern itself with the costs of changing from one series to another; i.e., from the B-52A to B-52B. Neither does it investigate specifically the costs to produce units after the first one or the relation of these later units to the first one produced. The term labor costs in this analysis refers to the number of direct labor manufacturing hours required to build the first aircraft produced and not the dollar value of these hours. This analysis examines seven aircraft produced by a single manufacturer. The general mission for which these aircraft were designed and their size are similar. Since this constitutes only a small segment of the total airframe manufacturing industry, it would be unwise to apply the results of this study indiscriminately without first verifying that the results apply to the aircraft in question.

Background

People in the business of cost estimating have long known that the estimates they provide are not likely to be

very accurate unless the cost estimate concerns something that has been done several times before. In fact, if they are estimating the cost of a new structure, particularly one that is not an extension of a previous structure, they realize that a three or four fold error in their estimate is not unlikely. Summers (Ref 37:12-15) cites several examples of cost estimates given for the construction of canals, railroads and nuclear power stations. Early estimates were off by a factor of from two to ten. The earlier in the program life the estimates were made, the poorer they tended to be. Original estimates are almost always on the low side. In examining 68 cost estimates on 22 major military programs, he found the estimates to be from 15% to 150% of actual program costs. 80% of the estimates examined were low. Actual costs are likely to be three times estimated costs if the estimate was made early in the program and the program was technologically difficult. The need to re-examine old estimating procedures and to search for new and better methods is clear.

Nearly all estimates for labor costs in the airframe industry center around the use of the log linear direct labor learning curve. There are several varieties of this curve and explanations of their differences can be found in many sources. The most frequently cited and detailed source is Asher (Ref 3:15-63). Another good source, though not so detailed is Brewer (Ref 7:43-66). The essential feature of all these curves is that the number of labor hours actually required to assemble an airframe follows a nearly straight line

when plotted on log log graph paper against the cumulative number of units produced. The reason why labor hours react in this manner is not clear. Nevertheless, equations can be developed, after the fact, that match almost perfectly the labor hours required for programs already completed These equations differ from program to program, but they are all of the same form:

$$Y = AX^b \tag{1}$$

where

Y	is the number of hours required to produce the Xth unit (or the number of hours per AMPR pound of airframe weight)
X	is the cumulative unit produced number
A	is a parameter whose value is equal to the number of hours (or hours per pound) to produce the first aircraft
b	is a parameter, unique to each individual program usually referred to as the slope

It should be clear that if the parameters A and b can be predicted accurately in advance of starting production, then labor costs could also be predicted very accurately. But predicting A and b is where the difficulty lies. It might be useful to consider a realistic example in order to develop some appreciation for the size of errors that might be introduced by only moderate prediction errors.

Suppose that the direct labor hours required for the first unit was estimated at 15 hours per pound of AMPR weight

prior to the start of production. Suppose also that the learning curve for the proposed aircraft was estimated to have a 74% slope. After production, it is observed that the first unit estimate was only 60% of the actual first unit costs. It is also observed that the slope estimate was exactly correct. These figures are ficticious, but they represent a very realistic description of what has often happened. In this case, actual labor costs to produce the first 100 aircraft would be 173% above the costs first estimated. But labor costs are a major cost element in the first 100 aircraft produced. So a first unit estimating error, that could be viewed as acceptable in the context of a historical examination of estimating errors, has led to a major error in the estimated costs for the first 100 aircraft produced.

The estimating method just described illustrates one method of estimating labor costs for a proposed program; i.e., estimating A (the hrs/lb for unit one) and b (the learning curve slope). An alternative method (Ref 13: IV-1, IV-2) would be to estimate the hours per pound for two or more units and then to solve algebraically for the slope. When this method is used, the points estimated are usually other than the first unit, so the parameter A must also be determined. It is also common to estimate the labor costs for unit 100 and the learning curve slope and then to solve for the parameter A (Ref 28:46-53). With both of these later methods, relatively minor errors in the estimates used can cause major errors in the resulting values of parameter A. No method is

presently available to put reasonable bounds on the values that could be assigned to A other than intuition.

The common way to apply intuition has been to examine historical records for aircraft of similar size and mission. The first unit labor costs from these historical records are adjusted, according to methods that vary from company to company, to produce an estimate for unit one of the new program. The chief cause of estimating errors is the small number of aircraft that are of similar size and mission to the one being proposed. Rational adjustments applied to only one or two previously produced aircraft can easily produce gross estimating errors. Air Force Negotiation Memorandum and Contract Proposals contain many examples of such estimates that were made in a seemingly rational way before the program was started, and that, after the fact, were greatly in error. One objective of this analysis is to provide some limits on what might be considered a good estimate of the parameter A.

The Problem

The problem to be investigated in this analysis, then, may be summarized in this way. What techniques and mathematical relations may be used to estimate the labor costs to produce the first airframe of a new type of aircraft? What reasonable bounds exist for such estimates? The objective of this research is to demonstrate that mathematical relations may be developed to accurately predict the values for the labor cost of the first airframe produced. Since this study uses data from only one manufacturer and on only one type of

aircraft, it is unlikely that the relations developed here will apply to the entire fixed wing airframe industry.

The Analytical Approach

The analytical approach presented in this study is flexible. It is not restricted by preconceived ideas as to the functional form resulting estimating equations must have. The form is determined by the data available for analysis and certain underlying assumptions. The chief assumption is that the labor costs are related in some identifiable way to some or all of the following factors:

1. Weight, speed, size, power, or some other engineering factor that is quantifiable.

2. Management and/or engineering competence.

3. Experience gained through the production of aircraft that are similar to the one being proposed.

4. Technology, in the sense that as time passes a given task becomes easier to do when new knowledge is applied to the performance of the task.

5. Variations in the way time and effort and materials are used in the research, development and planning that precede the production of a new aircraft.

It is also assumed that these factors may be arranged in some meaningful mathematical way to predict labor costs. As many combinations of these factors will be tested as the available data will allow.

One factor, that of management and engineering competence, must be abandoned at the outset. Since only one

manufacturer was willing to provide detailed information about his manufacturing methods, it is impossible to compare these methods to those of another manufacturer. Without such comparison, few valid conclusions can be made.

This study describes a search for a mathematical relation that will predict labor costs for aircraft of all sizes produced in any time period. The intuitive guide used in formulating and investigating various relationships can be stated as follows: Labor costs should be related in some way to the engineering complexity of the aircraft in question. Engineering complexity might be expressed mathematically as a function of weight, speed, thrust, payload, or other performance factors taken singly or in combination. When a suitable statement of engineering complexity is selected, it must be adjusted for time before comparisons can be made between aircraft that were not produced at the same time. The point here is that some mathematical function of time can be used to account for differences in the state of the art when aircraft are produced in different years. When a suitable combination of engineering complexity and time is selected and historical data is examined in relation to this selected expression, the labor costs for the production of certain aircraft will seem unusually high and others will seem unusually low. Where this trend is consistent, the variation might be reasonably explained by differences in the way resources were used in research and development effort for the aircraft in question, or in the way resources were used in the planning for the production of the aircraft.

These ideas provided the basis for gathering data and selecting variables for the graphic and mathematical analysis that follows. Before turning to these subjects, it might be well to outline some of the fundamental properties of the direct labor learning curve since much of the data used in this analysis was obtained directly from these curves.

II. Preliminary Relations

This study makes much use of a few symbols with specific and rather narrow meanings. Additionally, much of the data analyzed is obtained through constructing the direct labor learning curves for specific aircraft and extracting data from the equations for these learning curves. For these reasons, it is desirable to discuss briefly the fundamental mathematical relations of the learning curve and how learning curves are obtained.

The Learning Curve

It is a frequently observed fact in the airframe industry, and other industries as well, that the number of hours required to perform a task react in a very predictable way when the task is repeated many times. If it takes 10 hours to perform a task after it has been done several times and learning is occurring, then it will only take some fraction of 10 hours to do the task when it has been performed twice as many times. For example, if it takes 10 hours to do a task the tenth time it is performed, and learning is occurring at a 80% rate, then it will take only eight hours to perform the task on the 20th repetition and 6.4 hours the 40th time. The time required would continue to be reduced to 0.8 of its starting value each time the number of repetitions is doubled. This relationship can be expressed by the mathematical equation

$$Y = AX^b \tag{2}$$

where

	Y	is the number of hours required to perform the task
	X	is the number of times the task has been performed
	A	is the time to perform the task the first time (21 in this example)
	b	is the arithmetic slope of the learning curve equation (-.32 in this example)

If either the equation, or the actual observed values, for Y and X were to be plotted on regular rectangular coordinate graph paper the results would look like Fig. 1.

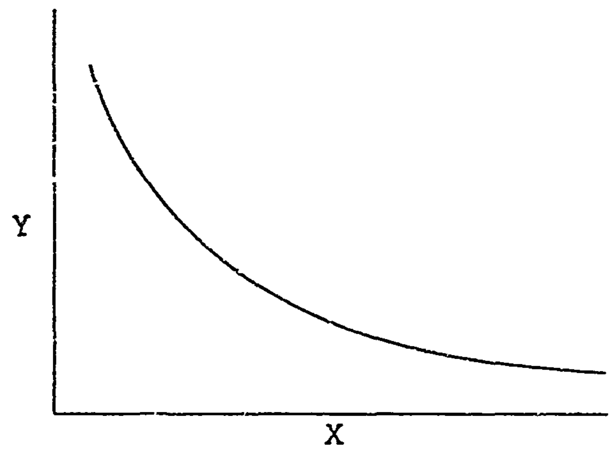

Fig. 1. $Y = AX^b$

If equation (2) or the actual observed values of Y and X were to be plotted on log log graph paper, the results would be quite different. The curve would appear as a straight line similar to the graph shown in Fig. 2. The same straight line can be produced in another way. If equation (2) is transformed by taking the logarithm of both sides, the resulting equation is

$$\text{Log}(Y) = \text{Log}(A) + b\text{Log}(X) \tag{3}$$

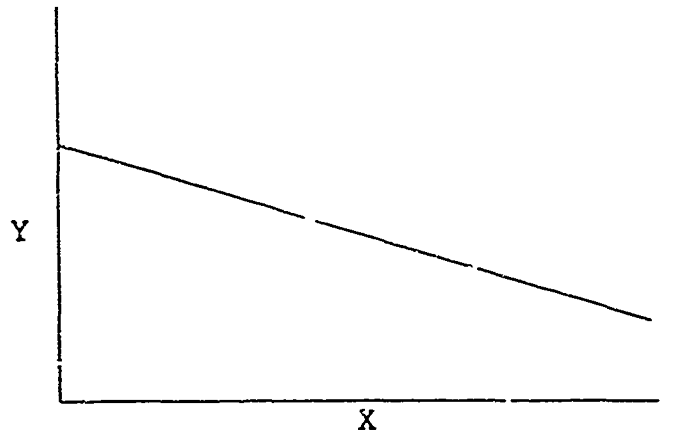

Fig. 2. $Y = AX^b$

(on log log graph paper)

If equation (3) is plotted on normal rectangular graph paper, the results would be the same as shown in Fig. 2. Plotting a multiplicative equation on log log paper then is equivalent to plotting the log of the equation on regular paper. Plotting on log log paper is very much simpler since it allows plotting the original Y and X values without looking up the logarithms first. If one is only concerned with the shape of the logarithmic graph, then log log paper is very useful.

The characteristic of the equation $Y = AX^b$ that makes it appear as a straight line on log log paper is the value of the parameter b. As long as it is a constant (i.e., it does not change with changing values of X) then the log log plot will be a straight line. If, when the observed values of Y and X are plotted on log log paper, the result is not a straight line, then the equation $Y = AX^b$ with b equal to a constant is not an appropriate model for the relationship being investigated.

In the airframe industry, the number of direct labor hours required to assemble an aircraft are not recorded for each aircraft that comes off the assembly line. Instead, the hours are recorded by lots of varying size. What must be done, if the learning curve is to be plotted, is to find the average number of hours required to assemble each aircraft in the lot (i.e., divide the total number of hours for the lot by the number of aircraft in the lot). The lot mid point must also be calculated. The desired value is the total number of aircraft produced up to and including the lot mid point. It is calculated in this manner. If the mid point for lot number 4 is desired and there were 30 aircraft produced before lot 4 was produced and there are 10 aircraft in lot 4, then the cumulative number for the mid point of lot 4 would be found by adding one half of the size of lot 4 (5) to the last cumulative number of the previous lot (30) for a cumulative lot mid point number of 35.

One further adjustment is usually made before the curve is plotted. When it is intended to compare one aircraft learning curve to another, the direct labor hours are usually divided by the AMPR weight of the aircraft being produced. It is this learning curve in hours per pound plotted against the cumulative units produced that is usually used when comparing two or more different aircraft.

Identifying the parameters of the equation that best describes the plotted learning curve is usually done by computer. The hours per pound and the corresponding cumulative

lot mid points are converted to their logarithmic values and a simple linear regression is then performed on these logarithmic values. The equations that are used to perform this regression are not relevant to this paper. They can be found in any good statistics or econometrics text book. Johnston (Ref 24:3-24) is an outstanding text as is Goldberger (Ref 18). Fisher (Ref 16:38-49) treats an example of logarithmic regression analysis in considerable detail.

The results of the regression will describe the parameters A and b in either one of two ways or both. If the results are from a simple logarithmic regression program, then the intercept coefficient will be Log(A) and the slope coefficient will be the arithmetic value of b. If the program used has been tailored for learning curve analysis, then A will be expressed in the actual value of hours per pound for the first unit assembled and b will be in a percentage describing the rate of learning. The relation between A and Log(A) can be found in a logarithm table. The relation between the percent of learning and the arithmetic slope (b) can be found from the equation:

$$\text{Log}\left(\frac{\%\ \text{learning rate}}{100}\right) = b \cdot \text{Log}(2) \tag{4}$$

One additional relation is worth pointing out. That is the relation between the costs to produce first unit and the parameter A of equation $Y = AX^b$. It has been noticed empirically that actual learning in the airframe industry can be closely matched after the fact by equations of this form.

The differences in equation values and actual observed values are usually minor, but these values are not the same. If $Y = AX^b$ is solved for the first unit labor costs per pound (i.e., $X = 1$) we find that $Y = A$ is this cost. This either may or may not be the actual labor cost for the first unit produced. Since labor costs are gathered by lot it would be nearly impossible to determine if they were the same or not. But it is really not important whether first unit labor costs is equal to A or not. We are really interested in the parameter A of the equation that best describes the labor costs of the whole program. The fact that this equation can be solved for an estimated value of first unit costs does not mean that actual first unit costs would be the value of A that would best predict labor costs for the whole program. The analysis that follows will be focused on ways to estimate, in advance of actual production, the values that A will most likely have. It is only in this sense that this study is an analysis of first unit labor costs.

III. The Data and Its Adjustment

A considerable amount of data in many forms was reviewed for appropriateness to this analysis. Most of the data was not in a form that would allow the analysis to be performed on it directly. It was necessary to make several adjustments in order to have the various programs comparable and to transform other information into a form that could be treated mathematically. The following paragraphs outline the data that was selected and how it was adjusted before the analysis was made. Figures that are provided have been altered to protect privileged information. They do, however, retain the essential characteristics of the actual data.

Direct Labor Data

Direct labor is one of the major expenditure classifications used in the airframe manufacturing industry. Other major expenditure classifications are Engineering hours, Tooling hours, Materials, Overhead and General Administration. There is no standard way to classify expenditures in the airframe industry. The classifications vary in name and content from firm to firm. They are similar however and no serious problem would be encountered in transforming the accounting classifications of two different firms into comparable figures. Since all the data used in this analysis was provided by one manufacturer, no alteration was necessary.

Since the data gathered is privileged information, its source and the aircraft they represent will not be

identified. The aircraft will be designated by arbitrary numbers, 1 to 7, assigned randomly.

The direct labor hour figures gathered represent those hours expended by workmen in machining, processing, fabricating and assembling the integral parts of the airframe structure. Asher (Ref 3:49) expands and treats this definition in considerable detail. Another classification is presented by Dei Rossi (Ref 13:II-2, II-6). The classification used by the manufacturer providing the data is consistent with these definitions. The data was available in both hour amounts and dollar amounts. Only hour amounts were collected to eliminate the requirement to consider changing wage and price rates.

As mentioned previously, the hour figures were not recorded by individual aircraft. Instead, they were recorded by consecutive groups of aircraft or lots. These lot sizes varied from two to 30 with the larger lot sizes usually occurring later in the production schedule. By dividing the total number of hours required to produce the lot by the lot size, the average number of hours per aircraft in the lot was obtained.

Information on all lots produced was not gathered. Since the analysis is concerned only with the characteristics of the first portion of the production run, unit 100 was selected as a cut off point. Lot information containing aircraft produced after unit 100 was not collected.

To make the hour figures for large aircraft comparable to the hour figures of smaller aircraft, all hour figures

were divided by the AMPR weight of the aircraft concerned. But the total AMPR weight of the aircraft could not be used because there was considerable variation in the amount of the aircraft that was actually produced and assembled in the plant. What was used was a figure called in-plant AMPR weight. It represents the AMPR weight of that portion of the aircraft that was actually produced and assembled in the plant. The resulting hours per in-plant AMPR pound then desecribes or approximates a figure that would have been realized if the entire aircraft had been produced and assembled in the plant.

There is one more series of adjustments necessary before the learning curves for the seven aircraft can be constructed and compared. Labor hour data for several of the aircraft began with two prototype aircraft. This was particularly true of the older programs. But prototype production is considerably different from the production of the first two aircraft of a production run. Prototypes are built in response to a company idea that a particular type of aircraft might be salable or in response to a request from a potential buyer. In either case prototype construction is essentially a one time construction of an aircraft to see if the design is feasible and salable. It is recognized that certain benefits may be realized that will make future production easier, but prototype construction may precede actual production by years and it is more nearly a one time construction project rather than the start of a production run.

Where prototype production was included in the direct labor data it was removed and counting of cumulative units began with the first aircraft of the normal production run. Average direct labor hours per in-plant AMPR pound per aircraft was calculated and regressed on the cumulative lot mid points to produce the estimated first unit labor cost designated Y_1.

It was felt that an alternative first unit labor cost reflecting what might have been learned during prototype construction would also be useful. There is no recognized method for adjusting learning curves for learning that occurs prior to the beginning of production so the following method was devised. Reasoning, as above, that the production of two prototypes does not produce the same learning as the production of the first two aircraft of a production run, it must produce learning equivalent to some fraction of two production aircraft. But since prototype programs vary in the amount of lead time before production is started, can be significantly different in design from the production aircraft, and can be produced at a different facility, it is not reasonable to assign the same fraction of two to all prototype programs. Considering these factors, and some others of a privileged nature, each prototype program was viewed separately and an intuitive judgment was made as to the amount of pre-production learning it could have produced.

All of the aircraft programs were then considered together and ranked according to the amount of pre-production

learning their prototype programs could have produced. For example, an aircraft program whose prototype was produced in response to a buyer request and where prototype construction and production occurred at different facilities would be ranked below a company conceived prototype produced in the production facility.

This ranking was then converted to a number between 0 and 1 and added to the cumulative lot mid points obtained earlier. Zero was assigned to those programs that did not have a prototype program. The higher numbers were assigned to those programs where a maximum of pre-production learning might have occurred. The earlier calculated average hours per pound per aircraft was then regressed on the new lot mid points. The first unit labor costs from these adjusted curves was designated Y_a.

A summary of the adjustments and the values of Y_1, Y_a, and Y_{a-1} are shown in Table I. The values have been altered to protect privileged information, but they still retain their essential relationships. The direct labor learning curves with prototype information included and no adjustments made are shown in Fig. 3. For comparison, Fig. 4 shows the learning curves after the prototype information has been removed and the above adjustments made. Notice that the curves in Fig. 4 have much more similarity in slope and intercept than those in Fig. 3. One convention will be adopted at this point. All future graphs, as well as Figs. 3 and 4, will be on log log coordinate graph paper unless specifically noted otherwise.

TABLE I

ESTIMATED FIRST UNIT LABOR COSTS

Aircraft	Y_1	Adjustment Factor	Y_a	Y_{a-1}
1	25.87	.75	28.49	28.49
2	21.93	.75	24.71	24.71
3	17.12	.75	19.77	18.75
4	19.00	.50	20.28	25.42
5	23.39	.25	24.45	24.65
6	20.22	0.00	20.22	20.90
7	27.31	0.00	27.31	27.31

Note: Symbols defined on page xi.

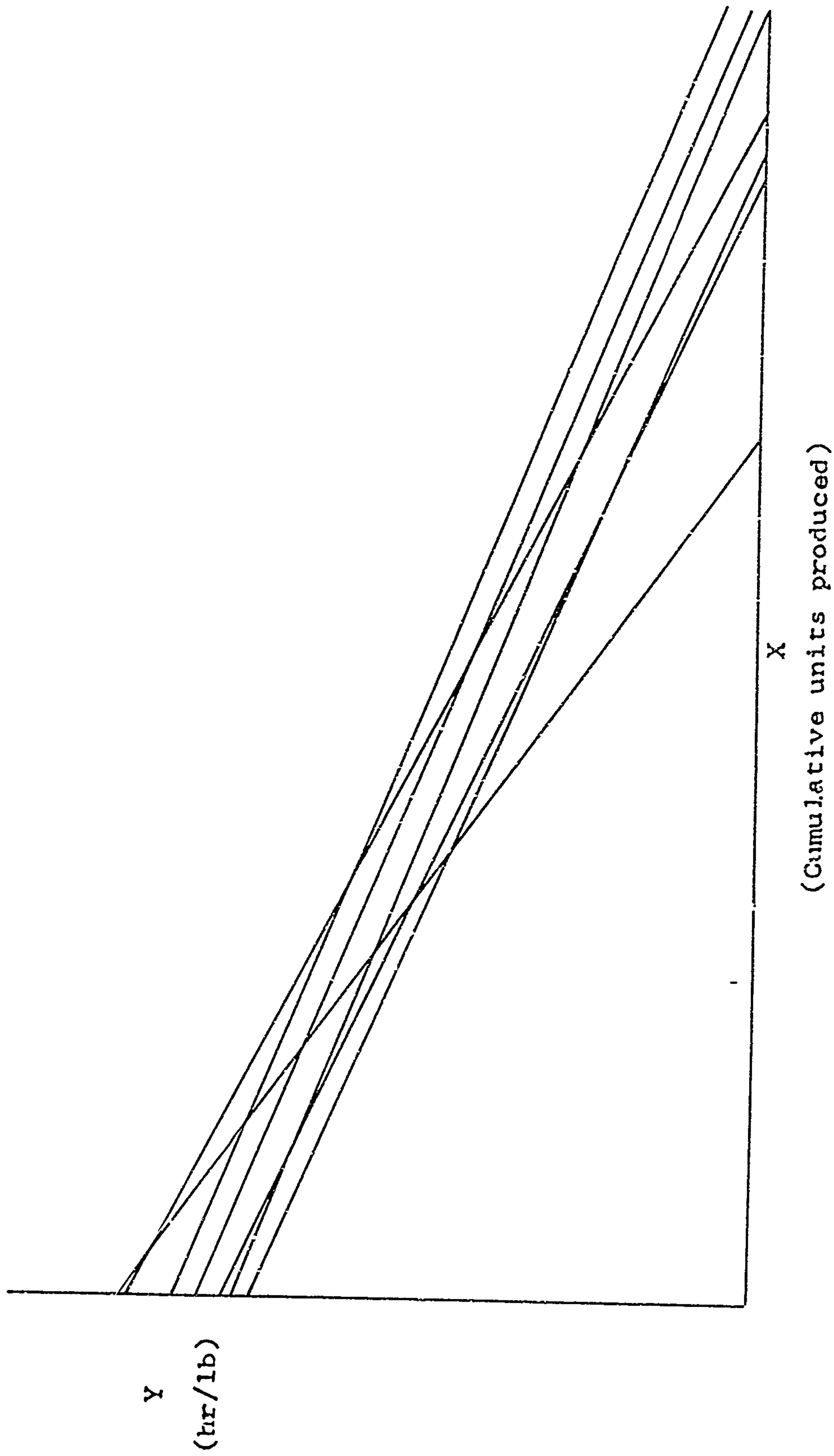

Fig. 3. Direct Labor Learning Curves
(from unadjusted raw data)

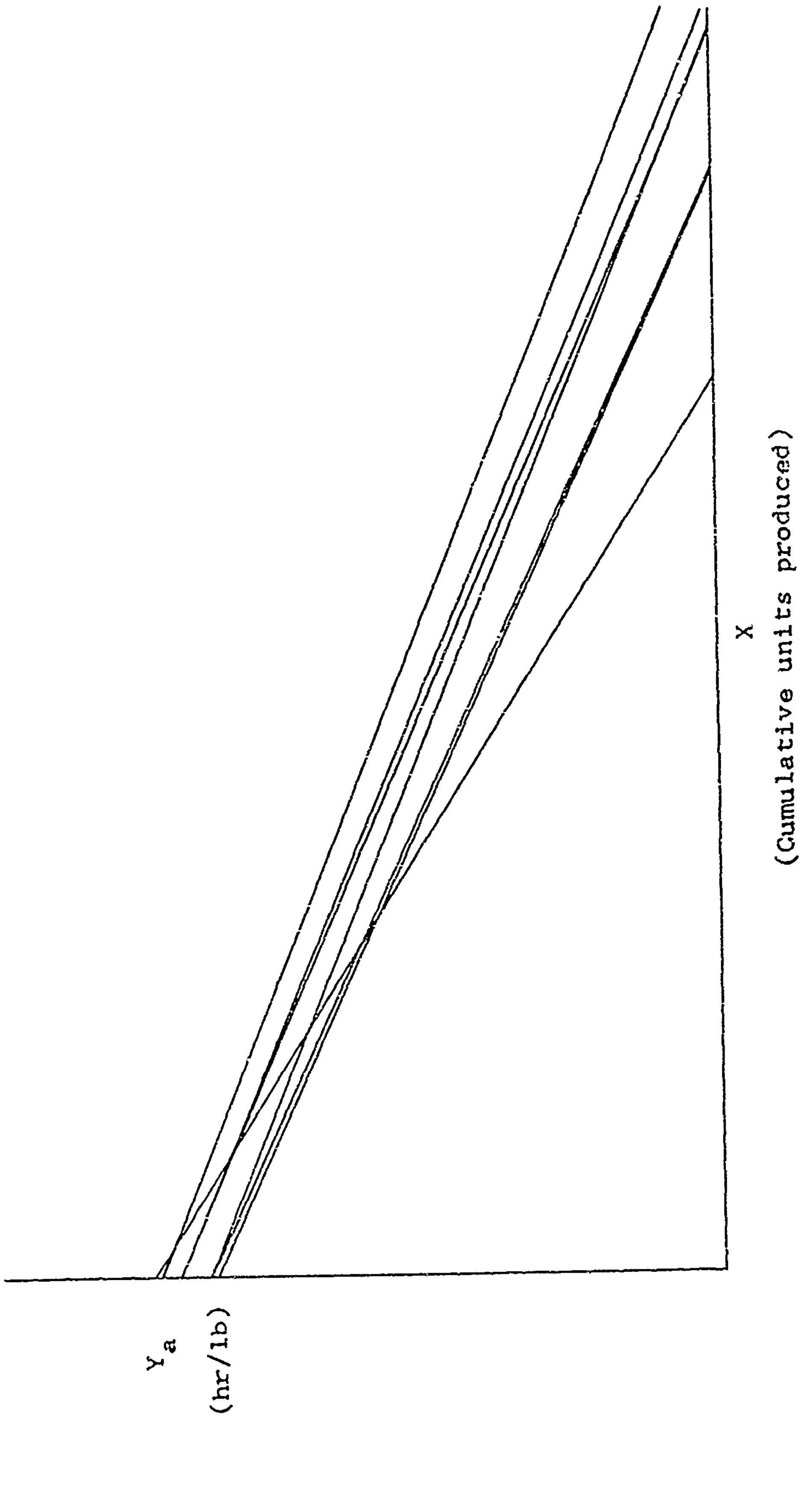

Fig. 4. Direct Labor Learning Curves
(after adjustment)

Two additional pieces of information are taken from the curves that produced Y_a. The first is the curve value for the 100th unit produced and it is designated Y_{100}. The second is found by extrapolating from the first lot mid point at its adjusted value, to the Y axis using the slope of the Y_a learning curve. This value could be considered a better approximation of what the actual first unit costs would be than either Y_1 or Y_a. This extrapolated value is designated Y_{a-1} and is also shown in Table I.

Performance Data

It was originally intended to gather as many variaoles as possible that could, in any way, be considered as a measure of engineering complexity for the aircraft in question. Data such as the number of design studies, the number of engineering drawings, the number of parts, airspeeds that reflect structural limitations at low and high altitude, power ratios and any others that might be available were sought. Some aircraft under consideration had all of these variables available. Most though, had only one or two. A choice had to be made between severely limiting the number of aircraft being studied or accepting only a few simple performance factors as measures of engineering difficulty. The latter choice was made.

Weight. Three weight figures were obtainable for all of the aircraft under consideration, maximum gross weight for take-off, AMPR weight and the difference between the two. The difference could be viewed as the amount of weight the AMPR airframe is designed to carry. Of the three weights,

the maximum gross weight for take-off is known earlier in the aircraft program life than the AMPR weight. Unless the AMPR weight or the weight difference proves to be a significantly better variable in estimating Y, maximum gross weight will be used in any equations that might be developed.

Speed. There were many speeds to choose from but few of these were arrived at in a consistent manner from aircraft to aircraft. There seemed to be an entirely unique set of speeds available for each aircraft, and the meaning of each set could only be understood in the context of its own program. It would have been necessary to have considerable time and the services of an aeronautical engineer to make the speeds comparable. Neither was available so the two that seemed most consistent were selected, maximum true airspeed at altitude, and maximum true air speed at sea level. Where these were in indicated air speed or miles per hour they were converted to knots true air speed. Later examination showed that the maximum speed at sea level could not be used to distinguish one aircraft from another in any meaningful way and it was discarded.

Time

The year and month was recorded when the first expenditure occurred in the manufacture of the first components of the first lot for each of the seven aircraft considered. This date will be used to represent the state of the technological art when construction of the aircraft began. However, calendar year without some modification is not suitable for

analytic purposes. Two types of modification will be used, exponential time and straight line time. The underlying idea in both of these approaches to the use of time is the same. If a task takes Y number of hours to do today, it can be expected that with technological advances it should take only some fraction of Y hours to do the same task ten years from now. There are some tasks that never change and, of course, this approach would not apply to them. But looking at an industry as a whole, or looking at the aggregation of many tasks, such a reduction in the required hours is bound to occur. The question of whether the saving is actually realized, or just diverted to newly created task will not be investigated directly. Time appears as a significant variable in some of the estimating relations developed in this study. Its importance and the role it plays will be examined after the relations are developed.

Exponential Time. If we are considering two points in time and the number of hours to do a task at each of these times, an exponential time approach would say that the number of hours required at the later time point is proportional to the time difference between the two points raised to a negative power between zero and one. That is:

$$Y_2 = Y_1 \cdot T^{exp} \qquad (5)$$

where:

Y_2 = the number of hours to do the task at the later time point

Y_1 = the number of hours to do the task at the earlier time point

T = the number of years between the two points

exp = a negative number between zero and one

An essential feature of exponential time is that its effect is not constant over different time intervals of the same length. This can be seen in Fig. 5 which approximates the curve of $T^{-.3}$. The points a, b, and c are spaced an equal distance apart. The base year is x_0. If it takes Y hours to do a task at time point a, Y will be reduced by a factor ΔY

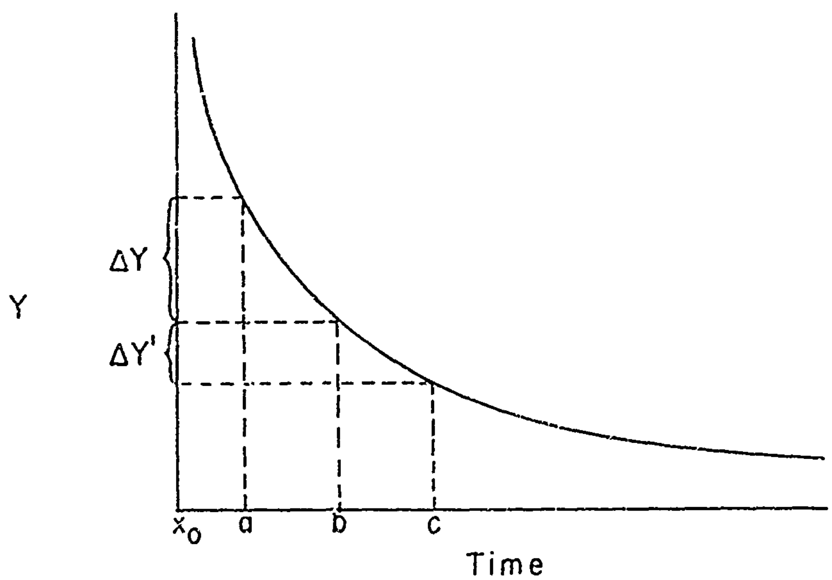

Fig. 5. $T^{-.3}$

when the task is performed at time point b. After another increment of time of the same length the task becomes reduced by an increment $\Delta Y'$. $\Delta Y'$ is smaller in absolute value and in percent of reduction than ΔY. If several events are considered together relative to one point in time, then the base year becomes very important since it will determine if the

events occur during the steep portion of the curve or the relatively flat portion. It should be clear that the largest reductions will occur in the first portion of the curve and relatively little adjustment is made in the later portion of the curve. If the events being considered do not relate in this way, then the choice of exponential time as adjustment factor would not be good.

Straight Line Time. An alternative to the exponential time concept that provides for a relatively constant reduction across all intervals is straight line time. A rate of reduction is selected that is to be applied every 10 years, say 75%. Then for every hour required in year zero, only .75 hours will be required ten years hence and $.75^2$ in twenty years. Time treated in this manner will be denoted $T_{.75}$ with the subscript denoting the rate of reduction. Straight line time is not actually a straight line, but it is so nearly so that no serious distortion is created in the 25 year time period covered by this analysis. Both $T^{-.25}$ and $T_{.75}$ are plotted in Fig. 6 across twenty-five years to show the difference in the way time reductions amounting to nearly the same quantity are applied. Table II shows several other pairs of exponential and straight line time for comparison.

Research and Development Data

Because of the privileged nature of this data, it will not be identified or discussed except to mention a few of its characteristics. This data was collected in four categories and when identification is necessary these

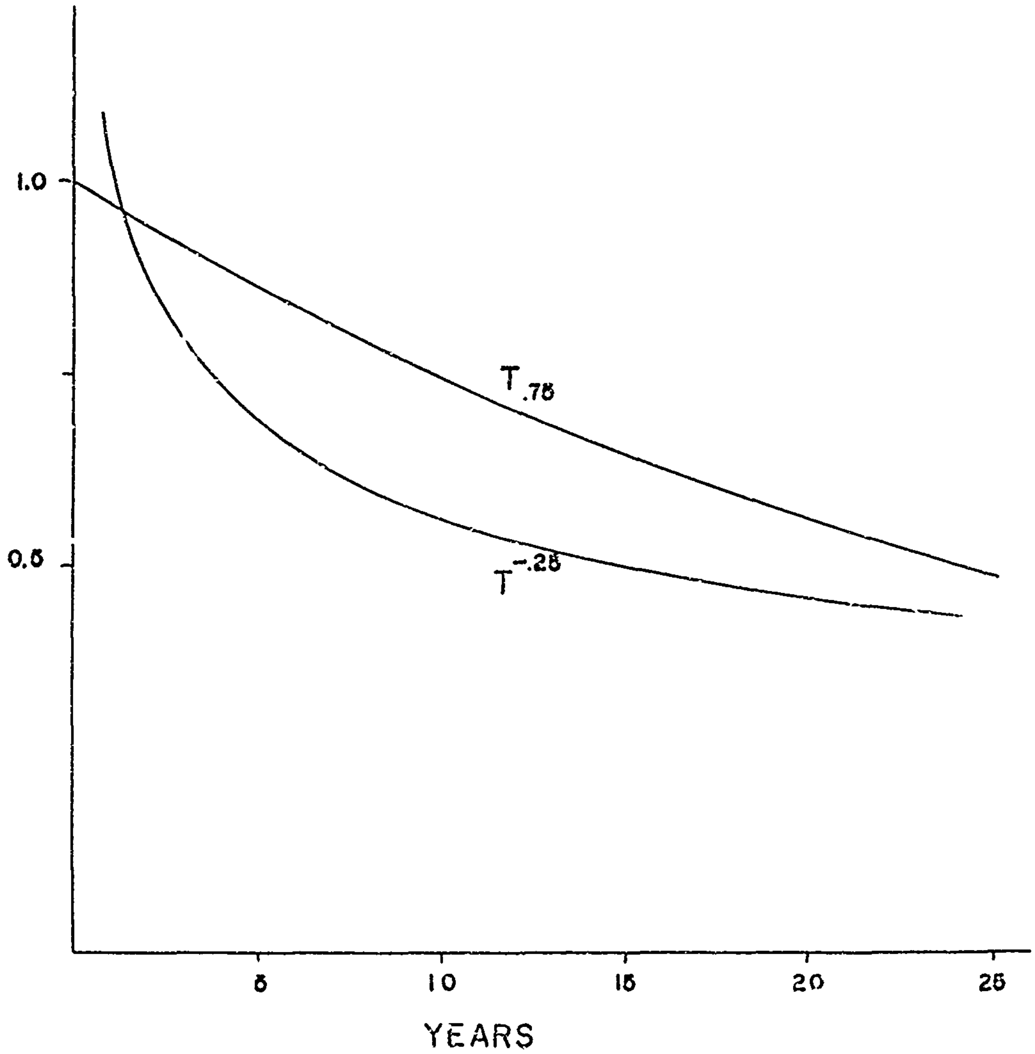

Fig. 6.

Adjustment Factor vs Time for $T_{.75}$ and $T^{-.25}$

(Rectangular Coordinates)

TABLE II

COMPARISON OF EXPONENTIAL AND STRAIGHT LINE TIME

Years	$T^{-.15}$	$T_{.85}$	$T^{-.25}$	$T_{.75}$	$T^{-.3}$	$T_{.65}$
5	.786	.922	.669	.866	.617	.806
10	.708	.850	.562	.75	.501	.65
15	.666	.784	.508	.649	.444	.524
20	.638	.723	.473	.563	.407	.423
25	.617	.666	.447	.487	.381	.341

categories will be referred to as R&D A through R&D D. They are related in a way that percentage figures are meaningful and they can be normalized by dividing by AMPR weight. In other words, R&D A per AMPR pound would have meaning and it could be plotted if desired.

Subcontracting Data

Data was gathered on the amount of subcontracting that occurred during the production of the first 100 of each of the aircraft considered. The trend was for the amount of subcontracting to increase in recent years. The amount of subcontracting varied considerably depending upon production rate, time constraints and how full or slack the aerospace industry was at the time of production. Company officials felt that subcontracting had both good points and bad. Where competent aerospace industries exist that specialize in one particular type of component, say auxillary power plants, the experience and expertise of such a company may allow them to do a job better and for less money than the prime contractor. However, where inexperienced contractors are involved, the experience is mixed. Their lack of experience often finds them over committed or performing a job for which they greatly underestimated the costs and complexities. In this case, things go badly. The overall impression is that if subcontracting is used wisely, it is at worse neutral and more probably a benefit.

This completes the discussion of the data gathered for this study and how it was modified for analysis. The

next chapter presents the analytical tools that will be used to perform the analysis.

IV. Analytical Methods

Three general mathematical tools will be used to analyze the data just described. They are graphic analysis, multiple regression analysis, and curve fitting by least squares. The purpose of this chapter is to explain how these tools will be used and what criteria will be used in judging the results.

In all three methods, the main test of any relationship developed will be its ability to predict. For every set of variables considered, there is at least one aircraft program whose labor cost presents a problem because the data representing the program appear out of agreement when viewed in context with the others. The aircraft that presents the problem varies depending on the particular set of variables being considered, and all aircraft become a problem candidate at one time or another. The existence of these problem points provides an ideal way to check the soundness of any estimating relation that might be developed.

Estimating relations are evaluated by many techniques most of which are related to some measure of how well the estimating relation reproduces the historical points upon which it is based. If estimating relations in this study were developed using all seven data points the resulting relations would produce better estimates of the problem points than if the relations were developed with the problem point excluded and only the six remaining points used. A relation developed without the problem point included in the data set may or may

not produce a good estimate of the problem point. If it produces a poor estimate, then there is good reason to question the validity of the relation even if it produces a fairly good estimate when the problem point is added to the data set and the relation is redeveloped.

The questions raised by the inability of a relation to accurately point toward a problem point are important. It could mean that the variables chosen for use in the estimating relation are inappropriate or that they are used in an inappropriate functional form. It could also mean that the aircraft represented by the problem point requires a different technology for its production and is not logically related to the other aircraft in the data set. It could mean that the labor costs for the problem point aircraft were adversely affected by some other factor that was unique to that aircraft program; i.e., some management policy, some engineering technique, or a major change in the program in mid stream, etc. A poor estimate of the problem point will not confirm the existence of any one or all of these possibilities. But a good estimate of the problem point allows consideration of these possibilities to be minimized. It also says that if the good estimating relation had been known and used before production of the problem aircraft had begun, it would have been of value. One would also have some faith in the ability of an estimating relation developed in this manner to accurately predict the labor costs of a new aircraft that represents an extension of technology past the technology of the problem point aircraft.

The term critical point will be used in this study to indicate the problem point chosen to test the ability of an estimating relation to predict. The aircraft representing the critical point will vary depending on the particular variables being used to develop the estimating relation. In general, it will be the point that appears most discontinuous when viewed in relation to the other six. All estimating relations will be developed with the critical point removed from the data set. Those relations that show the ability to estimate the critical well will be selected for refinement and further testing.

Graphic Analysis

Graphic analysis is the least sophisticated yet the most valuable method used because of its ability to improve understanding of how variables relate. Nearly every combination of variables possible was graphed but only those that have some special significance are included here. Those that show some important relationship are included along with some that show a complete absence of any recognizable relations. This is done to show why certain paths of analysis were pursued and others omitted.

Several conventions are adopted in the graphs to be presented. First, as mentioned previously, since the bulk of the graphs to be presented were constructed on log log paper, unless specifically noted otherwise, all graphs can be considered log log. Scales will be omitted from all graphs to protect privileged information. When referring to a specific

point on any graph, the points on that graph will be denoted by the letters a through g. The furthest point to the left will be designated as point a and the remainder be assigned consecutively to the right. Since there are usually seven points on each graph, the right most point would be point g.

The criteria for judging the worth of a graph is the reasonableness of an assumption that there exists a continuous, relatively smooth underlying function. Where this assumption seems plausible, and this underlying function helps to explain most of the points on the graph, then a search for that function will be pursued.

Multiple Regression Analysis

A computer program was designed to perform multiple regression analysis on equations of the form:

$$Y = B_o + B_1X_1 + B_2X_2 + \cdots\cdots + B_jX_j + E \qquad (6)$$

and

$$Y = e^{B_o} \cdot X_1^{B_1} \cdots\cdots X_j^{B_j} \cdot e^{E} \qquad (7)$$

The regression on equation (7) is performed by taking the logarithm of both sides which gives the linear equation:

$$\mathrm{Log}(Y) = B_o + B_1 \cdot \mathrm{Log}(X_1) + \cdots + B_j \cdot \mathrm{Log}(X_j) + E \qquad (8)$$

In the analysis performed, Y is usually hours or hours per pound and the X_i's are various combinations of performance data and time. E is the error term. The computer program solves for the values of the B_i's, the coefficient of deter-

mination (R^2), the partial coefficient of determination for each independent variable ($R_i^{\,2}$), an estimate of the variance of E and the F statistic.

Statistical tests will not be emphasized in the selection of usable estimating equations for two reasons. First, to use the F statistic, it requires the assumption of normality in the distribution of E with a mean of zero. There are few good reasons for making this assumption except to provide a statistical test. Second, and more important, is that the number of observations being regressed is not large enough to support statistical conclusions on more than one variable (Ref 26:60, 61).

The prime measures of validity will be the ability to predict as explained above, and the coefficients of determination (R^2) and the partial coefficients of determination ($R_i^{\,2}$). R^2 may be viewed as that percent of the variation of Y around its mean that is explained by the B_i's and the X_i's of the regression equation. A strong relationship would have a R^2 near 1.0 and a weak relationship R^2 would be near zero. The partial coefficients of determination ($R_i^{\,2}$) may be viewed as the percent of remaining variation that is explained by the addition of the ith variable to the regression equation after all other variables have been regressed. High values and a tendency toward equality of all the $R_i^{\,2}$'s would be desirable (Ref 18:199, 177; 25:31).

The equations used in the computer program and the linear algebra necessary to solve them would add only length

to this analysis and will therefore be omitted. The computer program, however, has been included as Appendix B. The above references provide de..iled explanations of the calculations and their interpretation as does Fisher (Ref 16: Section VI).

Curve Fitting

Where the requirement for linearity or log linearity cannot be met, or where the functional relationship between two or more variables is to be exactly specified, then curve fitting by least squares is used. For example, if a graph shows the relation between Y and X not to be a straight line, it might be desirable to derive an estimating equation in the form of a quadratic such as:

$$Y = B_0 + B_1X + B_2X^2 \tag{9}$$

The above computer program can be used to solve for the B_i's and the coefficient of determination. R^2 may then be used as a measure of goodness of fit (Ref 19:40).

In another situation it might be desirable to apply an adjustment factor, Z, to one of the variables and fit a quadratic in the logarithm of the resulting variables. An equation of this form would be:

$$\text{Log}(Y \cdot Z) = B_0 + B_1 \cdot \text{Log}(X) + B_2 \{\text{Log}(X)\}^2 \tag{10}$$

The solution for the value of the B_i's in this equation and any other where the exact relationship can be specified is a simple matter for the least squares curve fitting program. Its major limitation is the imagination of the user.

V. Analysis of the Data

The analysis of the labor cost data will proceed in three distinct phases. The first will be graphic. Hour labor data, both adjusted and unadjusted will be plotted against the various weight, speed, and time variables with the hope of identifying suitable variable combinations for the second phase, the mathematical analysis. Regression analysis will seek functional relationships where there appear to be linear or log linear relations between the variables. Where the graphic analysis suggests that there is a non-linear relation, assumptions will be made as to the likely form of the relation and curves will then be fitted to the data. The final phase of the analysis will be statistical. The best functions from the mathematical analysis will be selected and statistical tests will be applied to judge how strong a relation exists between the variables and how well the selected functions are likely to predict labor costs for future aircraft.

Graphic Analysis

When labor costs of two or more different aircraft are compared, the comparison is usually done in terms of direct labor hours per AMPR pound of airframe weight. Figs. 7, 8, and 9 plot this quantity against Time, Weight, and Maximum Airspeed respectively. Y is the general symbol adopted for hours per AMPR pound. Y_1 is the first unit labor costs found from the learning curve after prototype aircraft have been

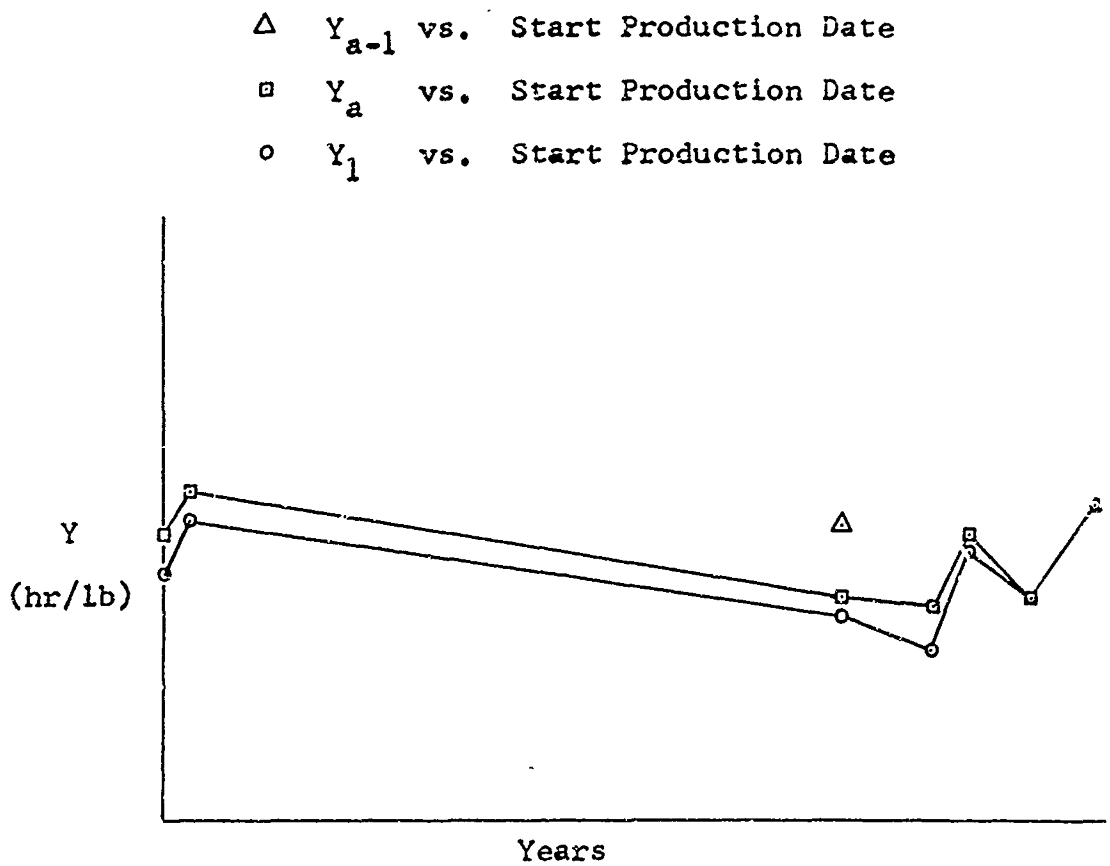

Fig. 7. First Unit Labor Costs
(in hours/AMPR pound) Vs Time

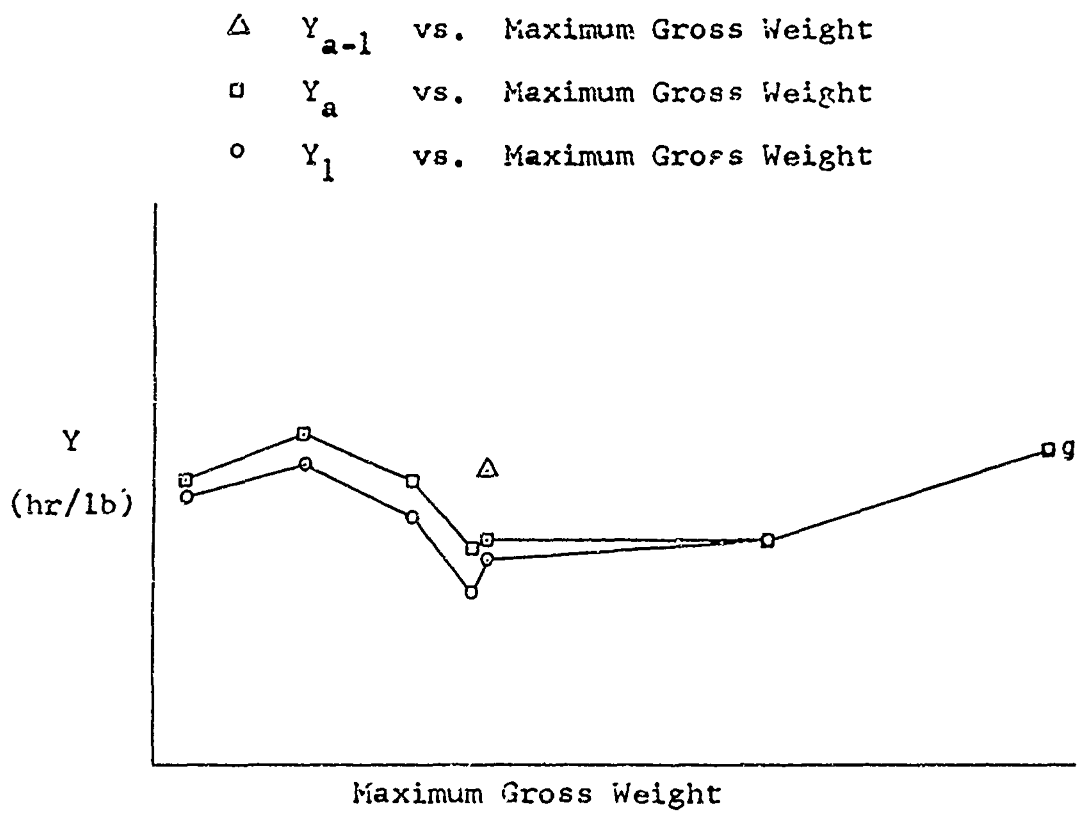

Fig. 8. First Unit Labor Costs
(in hours/AMPR pound) Vs Weight

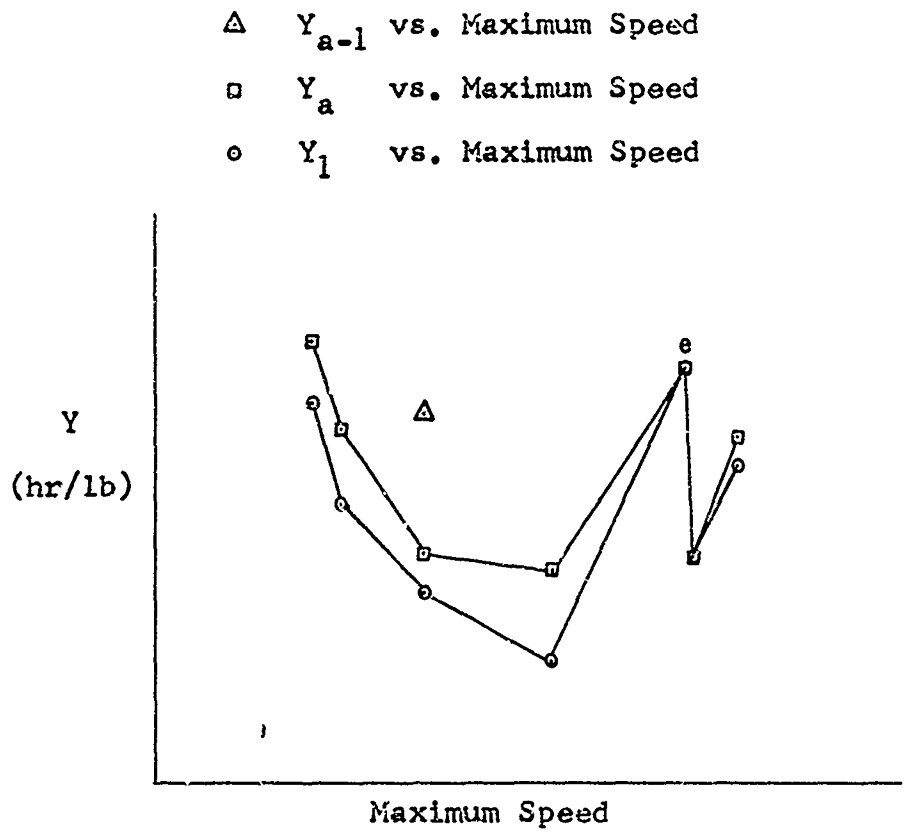

Fig. 9. First Unit Labor Costs
(in hours/AMPR pound) Vs Speed

removed. Y_a is Y_1 adjusted for pre-production learning. The graphs show Y_1 and Y_a. Also plotted on these first three figures is Y_{a-1} where it is significantly different from Y_a. Y_{a-1} is found by extrapolating from the first lot mid point to the Y axis using the slope of Y_a. Because Y_{a-1} is so similar to Y_a, except at one point, it will not be shown on further graphs.

The most important thing to note from these figures is that no one variable alone presents a reasonable explanation of labor costs. If a relation is to be identified, it will require two or more variables. It might be hypothesized that there is an underlying shallow U shaped curve in all three cases, or that both Y vs Time and Y vs Weight could be represented by a slightly downward sloping straight line. But neither assumption is clear or strong. These are possibilities that will be investigated further in the mathematical analysis.

Since Y is labor hours divided by AMPR weight, it might be instructive to see the picture of labor hours alone plotted against the same variables. To do this, Y_1 and Y_a were multiplied by their respective AMPR weights and designated H_1 and H_a. H_1 plotted against Time, Weight, and Speed are shown in Figs. 10, 11, and 12 respectively. In the case of time and speed, there is no recognizable relationship, but this is not so in the case of weight. Clearly, a straight line or a shallowly upward curving line could be underlying the H_1 vs Weight curve of Fig. 11.

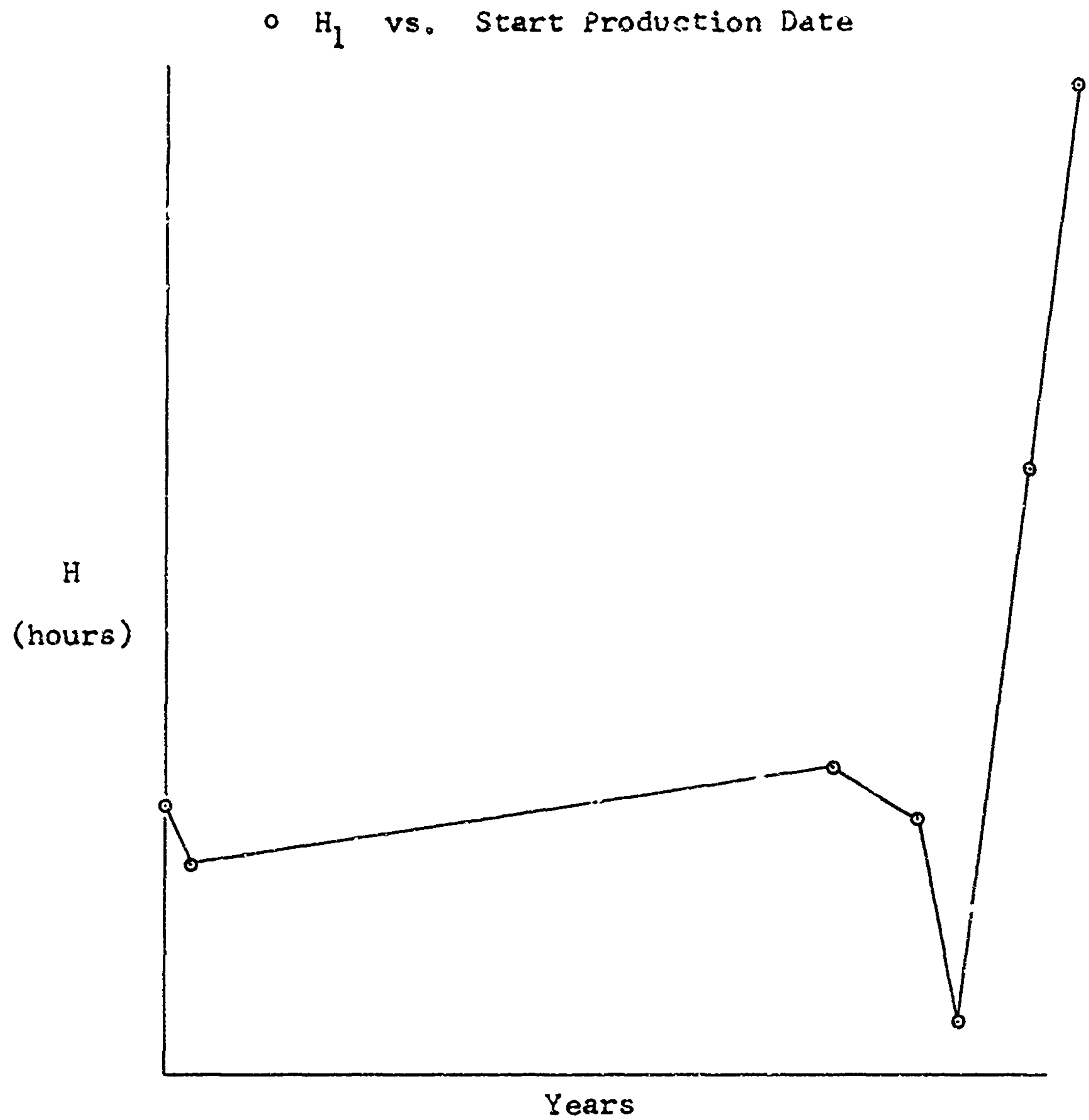

Fig. 10. First Unit Labor Costs
(in hours) Vs Time

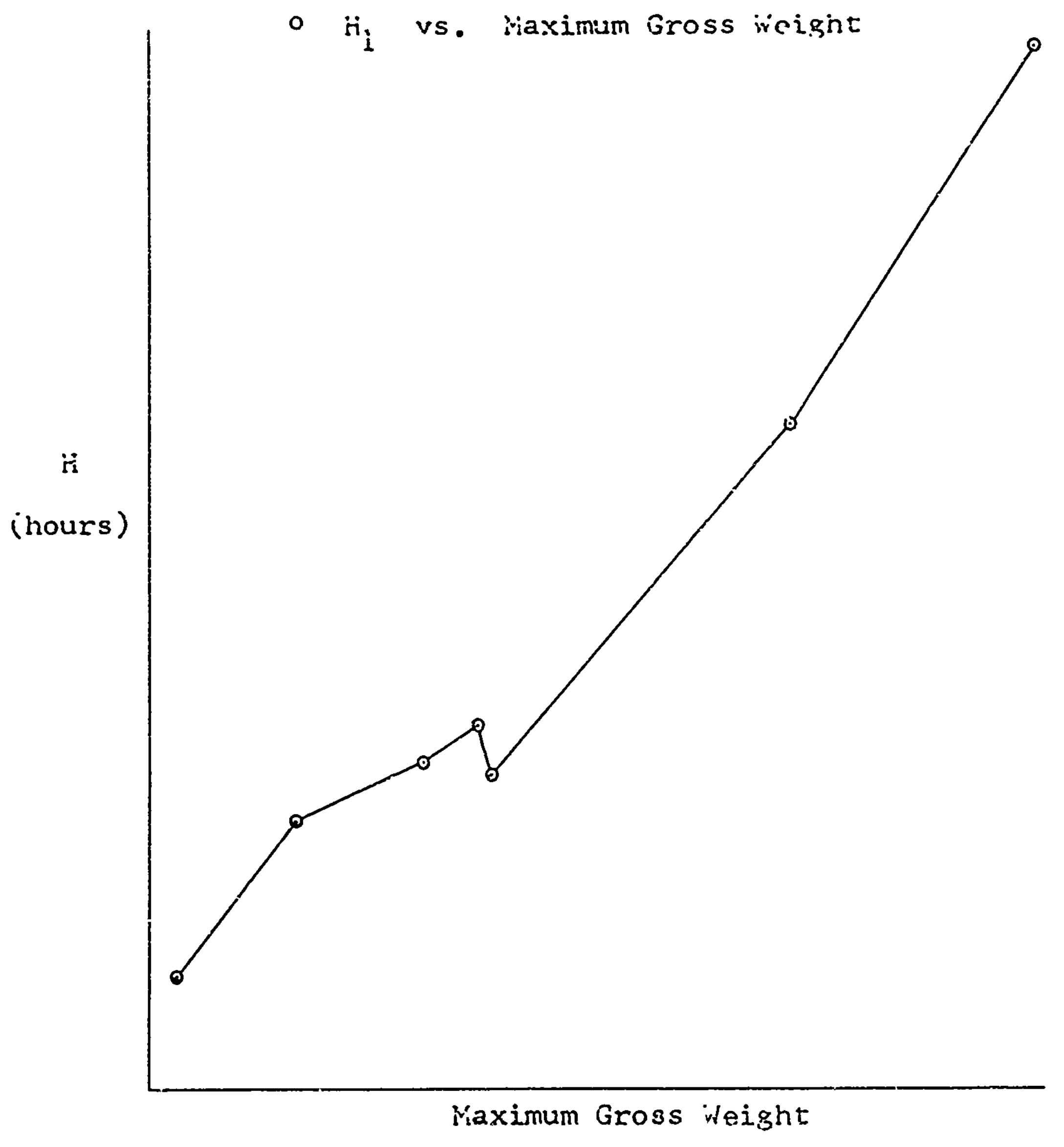

Fig. 11. First Unit Labor Costs
(in hours) Vs Weight

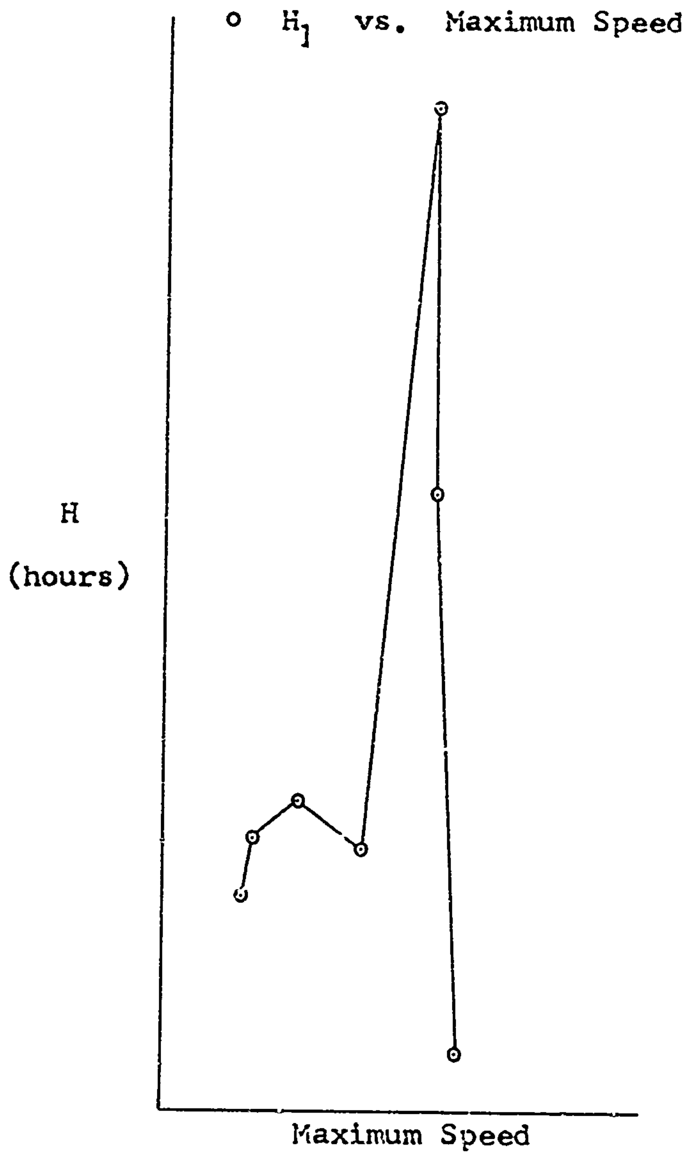

Fig. 12. First Unit Labor Costs (in hours) Vs Speed

These first six figures were all plotted on log log graph paper. Essentially the same picture is presented on rectangular coordinate paper and they are not shown for that reason. These six figures point the way for the mathematical analysis that follows, namely a search for linear and log linear combinations of two or more variables that explain Y and H, and a search for non-linear functions that explain the curve of H vs Weight.

Mathematic Analysis

Linear Regression. The first series of regressions seek to relate Y to speed, weight, and time as the independent variables. These independent variables will be taken two and three at a time by selecting one each from the following categories:

1. V
2. W_a, W_g, W_{g-a}
3. T^{exp}, $T_{.8}$

The equations formed will be of the following form:

$$Y = B_0 + B_1 \cdot V + B_2 \cdot W + B_3 \cdot T^{exp} + E \qquad (11)$$

or

$$\mathrm{Log}(Y) = B_0 + B_1 \cdot \mathrm{Log}(V) + B_2 \cdot \mathrm{Log}(W) + B_3 \cdot \mathrm{Log}(T^{exp}) + E \qquad (12)$$

The use of variables W_a and W_{g-a} produce nearly identical results in all of the regressions where they are used. For simplicity sake, W_{g-a} will be omitted from further consideration. W_g produced consistently slightly better results than W_a so where there are a considerable number of

combinations to be checked, only a few using W_a will be checked and reported.

Point g in Fig. 8, Y vs Weight, and Point e in Fig. 9, Y vs Speed, are the most discontinuous and therefore the most difficult to estimate. The initial regressions were performed using information on only six of the seven aircraft for which information is available. That is, these two critical points were excluded when either weight or speed was used as an independent variable. The regressions were then repeated with these critical points included. The results of these two series were then compared for their ability to accurately predict the critical points. Additional comparisons can be made from the coefficients of determination.

The results of these regressions are summarized in Table III. The table shows only a selected few of the variable combinations tested. Those selected were chosen to present a balanced view of the results which, as can be seen, were quite poor. The R^2 are quite low and the partial coefficients of determination (R_i^2) are low and unbalanced. All of the regressions using six observations produce large deviations in the resulting estimates of the critical points (denoted in the table by Y*-Y). The inclusion of the seventh observation, as expected, produced better estimates of the critical points but these better estimates were still quite poor. If normality in the error terms (E) of the regression equations is assumed and the relation between the variables tested with the F statistic (Ref 24:123), a significant rela-

TABLE III

SUMMARY OF REGRESSION ANALYSIS OF Y ON V, W, T^{exp}

Dependent Variable	Y_1	Y_a	Y_1	Y_a	Y_1	Y_a
Independent Variables	V W_g	V W_g	V W_g T^{exp}	V W_g T^{exp}	V W_g T^{exp}	V W_g T^{exp}
Equation Form	Log	Log	Log	Log	Linear	Linear
Sample Size	7	7	6	6	6	6
R^2	.023	.038	.669	.891	.703	.831
$R_1{}^2$	.00	.04	.44	.49	.59	.63
$R_2{}^2$	.02	.00	.11	.54	.03	.00
$R_3{}^2$	–	–	.55	.75	.65	.73
Y*-Y	-4.47	-4.0	-9.66	-9.52	-9.11	-11.40
F Ratio	.05	.08	1.5	5.4	1.6	3.3
F Statistic	4.32	4.32	9.16	9.16	9.16	9.16

Note: Symbols defined on pages ix, x, and xi.

tion between the variables is rejected in every case with a critical region of size .10.

The above regressions used T^{exp} in two forms. The first was calculated by selecting as a base year one year prior to the start of production for the oldest aircraft in the sample. This base year (and month) was then subtracted from the start production dates for the remaining aircraft to produce a time scale. The second form of T^{exp} used a base year five years prior to the first and was calculated in a similar manner. Both of these variables produce essentially the same results and the better of the two is shown in Table III.

The next series of regressions was designed to check if time used in another form produced better results. The concept of straight line time was substituted for exponential time and the series repeated. Straight line time is calculated and used in the following way.

A suitable learning fraction (r) must first be selected to represent the amount of reduction in labor requirements that might occur if a task were to be performed at one point in time rather than ten years earlier. Then the time adjustment factor for the start production date of aircraft i can be found by the equation:

$$T_{r,i} = r^{\left(\frac{d_i - d_b}{10}\right)} \qquad i = 1 \cdots 7 \tag{13}$$

where

r is the selected learning fraction between 0 and 1.0

d_i is the start production date in years for the ith aircraft (to the nearest twelfth)

d_b is the selected base year (to the nearest twelfth) and remains the same for all seven calculations

Equation (13) raises the learning fraction to a power equal to one tenth the difference between two dates. Tables IV and V show the resulting time adjustment factors for a learning fraction of 0.8 and base years of 1945 and 1970 respectively.

TABLE IV

TIME ADJUSTMENT FACTORS

$r = .8$ $d_b = 1945$

Start Production Date	1945	1950	1955	1960	1965	1970
Time Factor	1.00	.895	.800	.715	.640	.572

TABLE V

TIME ADJUSTMENT FACTORS

$r = .8$ $d_b = 1970$

Start Production Date	1945	1950	1955	1960	1965	1970
Time Adjustment Factor	1.75	1.57	1.40	1.25	1.12	1.00

Table IV can be interpreted to show the labor requirements if the performance of a 1945 task requiring one hour was delayed to the later years shown. Table V on the other hand shows the labor requirements if a task requiring one hour in 1970 was performed at an earlier date.

Both of these tables assume that there is a twenty percent reduction in labor requirements each ten years. From Table V it is possible to make estimates of present labor requirements if tasks performed in earlier years were actually performed for the first time in 1970 and 20% saving in labor is an appropriate figure. By dividing the actual historical labor expenditures incurred in 1945 by 1.75 the estimated 1970 labor requirements for a 1945 task can be found.

Using equation (13) and d_b equal to the start production date for the last aircraft produced in the study sample, straight line time adjustment factors were calculated for all aircraft in the sample for each .05 increment between .65 and .90. Y_1 and Y_a were divided by these time adjustment factors and the resulting Y/T_r's were regressed on the same weight and speed variables used above. The results were very similar to the first series of regressions. Table VI shows the results that were obtained for some variable combinations not shown in Table III. The reason for showing only $T_{.8}$ is not because $T_{.8}$ produced any better or worse results than other straight line time fractions, but that $T_{.8}$ will play an important role in later analysis and this information will be available for comparison.

The graph of total labor hours vs weight shown in Fig. 11 provides the motivation for the next series of regressions. The curve of H vs Weight is so much more continuous in appearance that it suggests H might be easier to estimate than Y. To determine if this is so, linear and log linear regressions

TABLE VI

SUMMARY OF REGRESSION OF $Y/T_{.8}$ ON V, W

Dependent Variable	$Y_1/T_{.8}$	$Y_1/T_{.8}$	$Y_a/T_{.8}$	$Y_a/T_{.8}$	$Y_1/T_{.8}$	$Y_a/T_{.8}$
Independent Variable	W_a	W_g	W_a	W_g	V	V
Equation Form	Log	Log	Log	Log	Log	Log
Sample Size	7	7	7	7	7	7
R^2	.381	.348	.338	.307	.494	.445
Y*-Y	-4.77	-5.25	-4.63	-5.03	-7.44	-6.75
F Ratio	3.07	2.67	2.56	2.21	4.89	4.01
F Statistic	4.06	4.06	4.06	4.06	4.06	4.06

TABLE VI Continued

Dependent Variable	$Y_1/T_{.8}$	$Y_1/T_{.8}$	$Y_a/T_{.8}$	$Y_a/T_{.8}$	$Y_1/T_{.8}$	$Y_a/T_{.8}$
Independent Variable	W_a	W_g	W_a	W_g	V	V
Equation Form	Linear	Linear	Linear	Linear	Linear	Linear
Sample Size	6	6	6	6	7	7
R^2	.040	.047	.000	.000	.472	.427
Y*-Y	-7.76	-7.99	-10.13	-6.75	-6.90	-6.32
F Ratio	.16	.19	.00	.00	4.47	3.73
F Statistic	4.54	4.54	4.54	4.54	4.06	4.06

were performed using the following variables. The dependent variables were chosen from:

$$H_1,\ H_a,\ \frac{H_1}{T_{.8}},\ \frac{H_a}{T_{.8}}\ .$$

The independent variables were chosen from:

1. V
2. W_g, W_a
3. T^{exp}

Where H/T was chosen as a dependent variable, only one or two of the variables from categories 1 and/or 2 were selected. Where H, or H_a, was selected as the dependent variable as many as three were selected with no more than one W variable in any one equation. The same two time scales for T^{exp} that were used in the first series of regressions were used in this series. Each set was regressed both with and without the critical points.

The results of this series of regression is shown in Table VII. From a statistical point of view the results are very much better than the previous series. However, the ability of these equations to estimate the critical points is again very poor. As in the previous tables, Table VII gives a cross section of the results and all of the combinations shown in the table were regressed without the critical points included. The results with the critical points included differed very little from the results shown in Table VII.

TABLE VII

SUMMARY OF REGRESSION OF H ON V, W, T

Dependent Variable	H_1	H_1	H_1	H_a	H_a	$H_1/T^{.8}$	$H_1/T^{.8}$	$H_a/T^{.8}$	$H_a/T^{.8}$
Independent Variable	V W_g T^{exp}	V W_g T^{exp}	V W_g	V W_g	V W_g	V W_a	V W_g	V W_a	V W_g
Equation Form	Log	Linear	Log	Log	Linear	Log	Linear	Log	Linear
R^2	.967	.991	.952	.971	.992	.947	.989	.988	.995
R_1^2	.183	.305	.024	.208	.409	.04	.33	.63	.36
R_2^2	.963	.987	.951	.971	.99	.93	.98	.98	.99
R_3^2	.303	.459	–	–	–	–	–	–	–
Y*–Y	–12.63	–9.32	–13.8	–13.21	–9.74	–12.36	–8.86	–11.21	–11.08
F Ratio	19.3	79.9	29.7	50.5	196.5	26.5	131.0	123.	231.8

Equations derived with the critical included were, as expected, better able to estimate the critical points but the improvement was at the expense of poorer estimates of other sample points. Overall, as estimating equations, they have high variance in their estimates of the sample points (±3.0 being typical).

One convention is adopted at this point. When equations are developed that estimate H rather than Y, the point being estimated will be divided by its AMPR weight and deviations expressed in terms of hours per AMPR pound. In this way, all deviations can be kept consistent and comparable.

The reason for the poor performance of these estimating equations can be seen by examining the partial coefficients of determination. The partial coefficients for weight variable is greater than .90 in every case and the partials for the other variables are low and erratic. These equations are essentially functions of weight alone with just minor influences by time and speed. A more balanced relation is needed for a good estimating relation.

Before continuing the search for a balanced relation, it might be well to point out some tentative conclusions based on the first three series of regressions. Linear and log linear combinations of weight, speed, and time apparently cannot provide low variance estimates for Y and H by themselves. Either Y and H are not estimable, or other variables must be added to these to explain Y and H, or an altogether different set of variables must be used if linear or log linear relations are to be found. The addition of more

variables to the estimating equations is not desirable. Adding more variables cannot make an equation produce poorer estimates. In fact, they can only stay as good or improve. By adding enough arbitrarily selected variables, an equation can be made to produce near perfect estimates where, in fact, the relations used may be nonsense. With the sample size used in this study, statistical inference on more than two variables would be extremely weak (Ref 26:61). Since other variables are not available, the search for estimating relations must be ended or turned to non-linear forms. It is toward non-linear forms that this study now turns.

Regression of Non-Linear Variables. Fig. 11 again provides the clues for the analysis that follows. In examining the individual aircraft in Fig. 11 (First Unit Labor Costs in Hours Vs Maximum Gross Weight) for likely reasons why their total hours deviated from some imaginary underlying curve, it became apparent that the deviations were related very strongly to time. Rough calculations showed that the older the start production date, the further the point was from an imaginary smooth curve. To see how strong the influence of time was, each H was divided by $T_{.75}$ and plotted on another graph. It was apparent that $T_{.75}$ was too large an adjustment and $T_{.8}$ was selected and plotted. The results of this adjustment are shown in Fig. 13, H_a and $H_a/T_{.8}$ vs AMPR weight. Compared to previous curves $H_a/T_{.8}$ vs AMPR weight is quite smooth and continuous. It strongly resembles the first quadrant graph of a parabola with its axis parallel to the

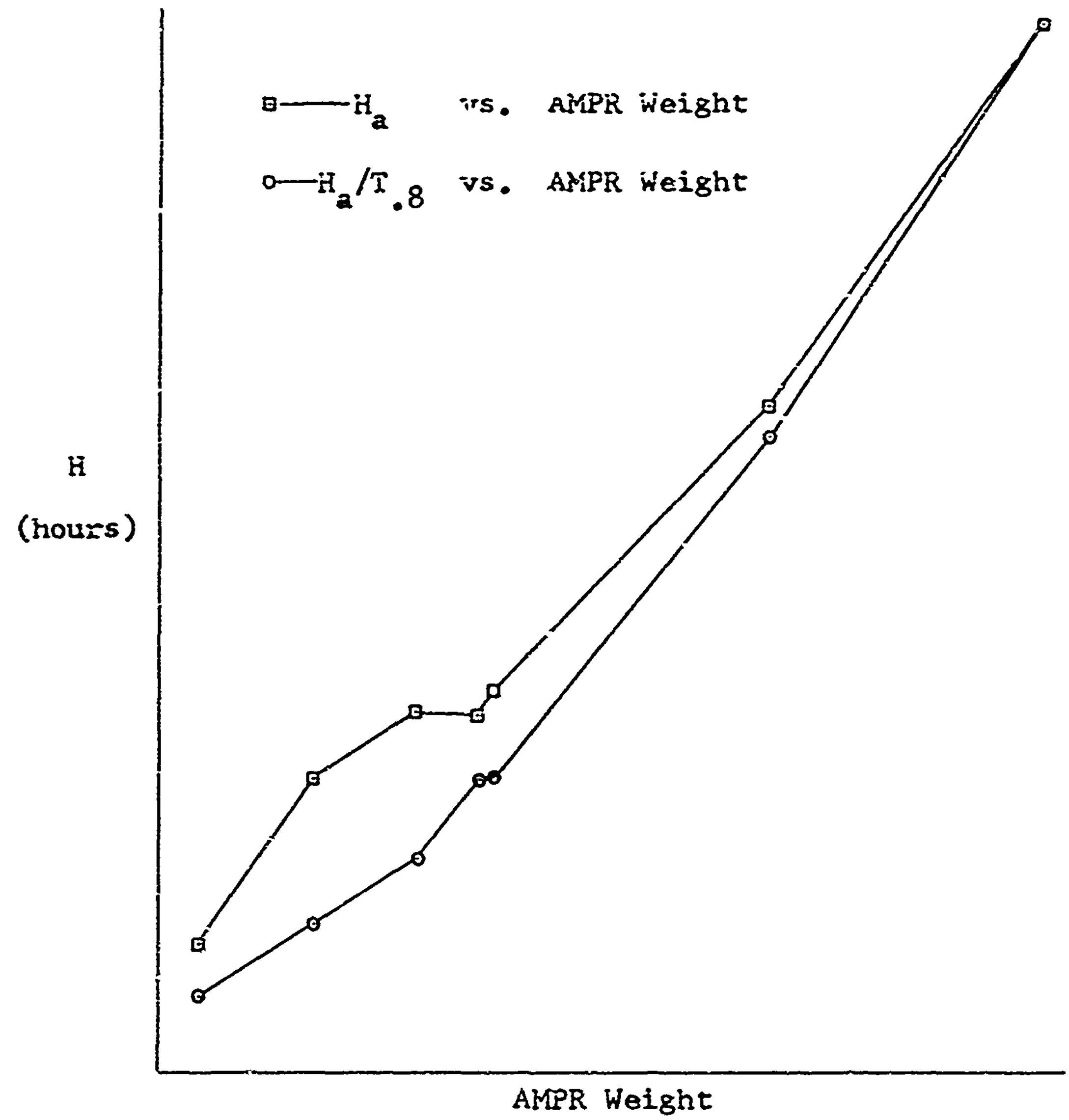

Fig. 13. First Unit Labor Costs
(in hours) Adjusted for Time Vs AMPR Weight

vertical axis. The graphs of $H_1/T_{.8}$ plotted against AMPR weight and both $H_1/T_{.8}$ and $H_a/T_{.8}$ plotted against gross weight are very similar. The general shape of these curves is retained on rectangular coordinate graph paper. The pictures presented by these graphs suggest that a quadratic equation in H/T and W and a quadratic equation in the logs of these variables might produce very good estimating equations. These equations would be of the form:

$$H/T_{.8} = B_o + B_1 \cdot W + B_2 \cdot (W)^2 + E \qquad (14a)$$

and

$$\mathrm{Log}(H/T_{.8}) = B_o + B_1 \cdot \mathrm{Log}(W) + B_2 \cdot \mathrm{Log}(W)^2 + E \qquad (14b)$$

Combinations of H_1, H_a, W_g, W_a were examined. Before showing the results, it might be well to perform the same type of time adjustment on Y vs Weight and examine the results for possible relations.

Fig. 14 shows the graph of $Y/T_{.8}$ vs AMPR Weight. The graph of Y vs Gross Weight is omitted for simplicity sake. This graph also suggests that a quadratic equation of the above form with H replaced by Y might give good estimating equations.

The variable combinations just mentioned give 16 possible quadratic estimating equations. Since the square of weight and the square of the log of weight can be calculated and used as variables, equations (14a) and (14b) are still suitable for linear regression. It is possible then to use a least squares linear regression program to fit the quadratic

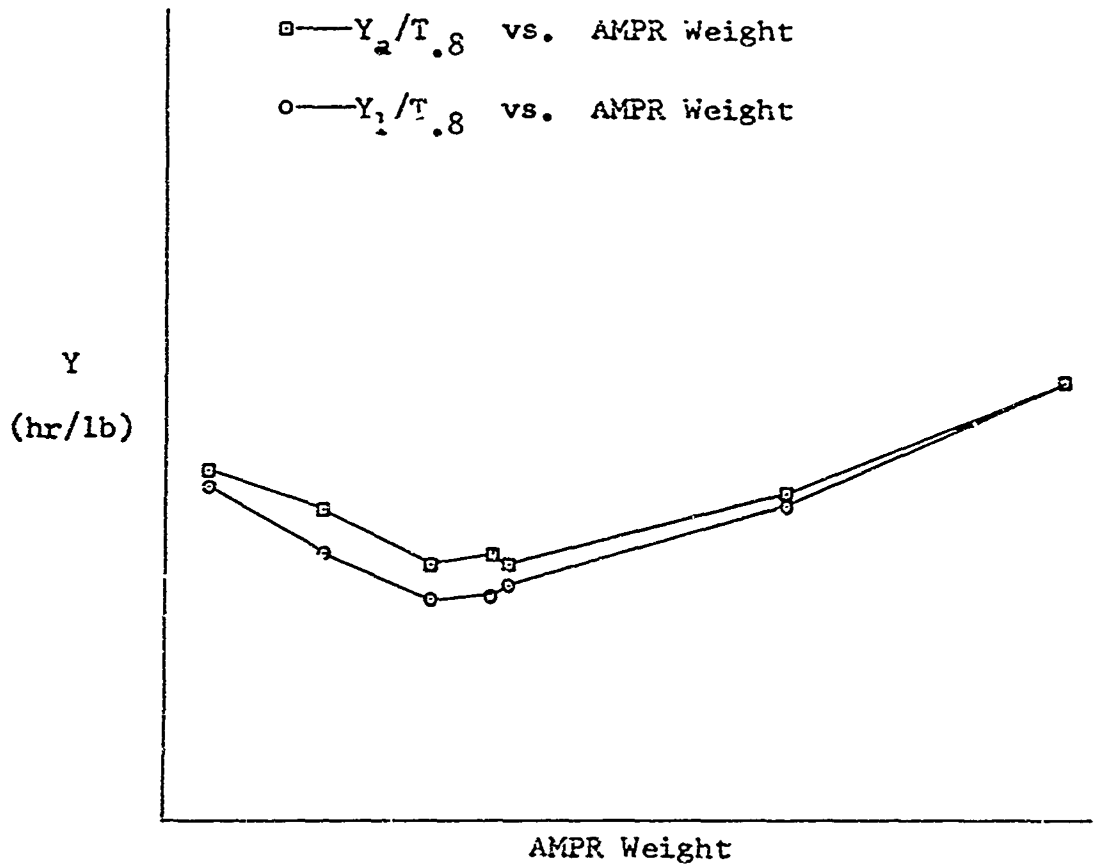

Fig. 14. First Unit Labor Costs (in hours/AMPR pound) Adjusted for Time Vs AMPR Weight

curves and have, at the same time, all the statistical tools that were available for multiple linear regression (Ref 19: 40, 107-123).

The regressions of these quadratic equations were performed in the same manner as the earlier series of regressions, that is, omitting the critical points and using a sample size of six and then repeating the regressions with the critical points included.

The results of these regressions are very good from a statistical point of view for both the sample size of six and seven. All of the equations derived from a sample size of seven produced low variance estimates of the critical points. However, that was not the case with the sample size of six. In fact, some of the estimates were the poorest of all the equations examined so far. In order to show the dangers of indiscriminately applying equations of this form, the results of all the quadratic regressions with the critical points excluded will be shown in the following tables. From the best of these, equations will be selected for statistical analysis.

The equations in Table VIII produced the poorest statistical results and had the greatest deviation in the estimates for the critical points. The critical point estimates of these equations were in error by as much as a factor of 5. In Tables IX, X, and XI the variable combination $H_1/T_{.8}$, or $Y_1/T_{.8}$ with W_a produced the poorest results. The remaining equations, with one exception, were all within 30% of the actual value of the critical point. These eight equations,

TABLE VIII

SUMMARY OF QUADRATIC REGRESSION $Y/T_{.8}$ ON W
(In Form of Equation (14a) Critical Points Excluded)

Equation Number	15	16	17	18
Independent Variable	$Y_1/T_{.8}$	$Y_1/T_{.8}$	$Y_a/T_{.8}$	$Y_a/T_{.8}$
Dependent Variable	W_a	W_g	W_a	W_g
R^2	.878	.891	.885	.879
$R_1{}^2$	.86	.87	.88	.87
$R_2{}^2$	.87	.88	.88	.88
Y*-Y	+121.5	+91.3	+98.9	+73.2
F Ratio	10.8	12.3	11.6	10.9
F Statistic	5.46	5.46	5.46	5.46

TABLE IX

SUMMARY OF QUADRATIC REGRESSION OF $H/T_{.8}$ ON W
(In Form of Equation (14a) Critical Points Excluded)

Equation Number	19	20	21	22
Independent Variable	$H_1/T_{.8}$	$H_1/T_{.8}$	$H_a/T_{.8}$	$H_a/T_{.8}$
Dependent Variable	W_a	W_g	W_a	W_g
R^2	.999	.998	.999	.999
$R_1{}^2$	.76	.59	.91	.89
$R_2{}^2$	.98	.87	.95	.87
Y*-Y	+9.03	+1.00	+4.20	+1.28
F Ratio	3788.	717.	3162.	1493.
F Statistic	5.46	5.46	5.46	5.46

TABLE X

SUMMARY OF LOG QUADRATIC REGRESSIO' 'T$_{.8}$ ON W
(In Form of Equation (14b) Critical Poi. s Excluded)

Equation Number	23	24	25	26
Independent Variable	$Y_1/T_{.8}$	$Y_1/T_{.8}$	$Y_a/T_{.8}$	$Y_a/T_{.8}$
Dependent Variable	W_a	W_g	W_a	W_g
R^2	.975	.980	.961	.963
$R_1{}^2$	.97	.98	.96	.98
$R_2{}^2$	.97	.98	.96	.98
Y*-Y	17.28	8.98	6.66	2.33
F Ratio	58.0	73.4	36.7	39.3
F Statistic	5.46	5.46	5.46	5.46

TABLE XI

SUMMARY OF LOG QUADRATIC REGRESSION H/T$_{.8}$ ON W
(In Form of Equation (14b) Critical Points Excluded)

Equation Number	27	28	29	30
Independent Variable	$H_1/T_{.8}$	$H_1/T_{.8}$	$H_a/T_{.8}$	$H_a/T_{.8}$
Dependent Variable	W_a	W_g	W_a	W_g
R^2	.999	.993	.999	.997
$R_1{}^2$	.94	.80	.84	.78
$R_2{}^2$	.97	.90	.96	.92
Y*-Y	+17.27	+3.89	6.65	-3.53
F Ratio	1058.	228.	1071.	480.
F Statistic	5.46	5.46	5.46	5.46

numbers 20, 21, 22, 25, 26, 28, 29, and 30 are selected for further analysis.

Before turning to the statistical tests, it is worth pointing out some relations between the variables that can be seen in the four tables. Generally, $Y_a/T_{.8}$ produces better estimates of the critical point than does $Y_1/T_{.8}$. The same is true of $H_a/T_{.8}$ over $H_1/T_{.8}$. H in general is better than Y and W_g is better than W_a. These trends were observed earlier in the linear regressions, but the linear equations were so poor that not much attention was paid to them. If these trends continues with the addition of the critical points to the sample size, then we should expect that equations 22 and 30 would produce the best results.

Statistical Analysis

The statistical analysis will be concerned with two areas. The first is the nature of the distribution of error terms, or residuals as they are often called, of the equations that are selected as good estimating equations. After determining an appropriate model for the error terms, the selected distribution will be used to provide a 80% confidence interval for expected value of the labor costs of a hypothetical 500,000 pound maximum gross weight aircraft produced in mid 1970.

The criteria used to select the good estimating equations must insure that they are statistically sound and capable of providing realistic estimates. To insure this, the equation regressed without critical points must possess a

coefficient of determination greater than .80 and be able to estimate the critical points within plus or minus 30% before it will be considered for the statistical analysis. All equations regressed were examined using this criteria and only the eight previously mentioned quadratic equations qualified.

Distribution of the Error Term. Table XII shows the deviations between the equation estimate for the hours per AMPR pound for each of the seven aircraft and the actual value of Y_1 or Y_a that the equation is based on. Where the equation is in log form or the dependent variable is in hours rather than hours per AMPR pound, these values were converted to hours per AMPR pound for calculating the deviations. These values are shown to provide some feel for the accuracy of the selected equations and are not the deviations to be analyzed for the distribution of the error terms. Notice that equations 20-22 are rather poor in their ability to estimate the first five aircraft when compared to the remaining equations, but they are very good on aircraft 6 and 7. The log quadratic equations on the other hand seem to be equally good throughout all aircraft.

Returning to the question of the distribution of the error terms, regression theory provides many ways to perform calculations of the range of possible errors of a prediction provided the assumption can be made that the error terms are distributed normally. To make use of these methods, it is necessary first to insure that the normal distribution is a

TABLE XII

SUMMARY OF DEVIATIONS

Equation Number Form*	20 Q	21 Q	22 Q	25 L	26 L	28 L	29 L	30 L
Quantity Estimated	Y_1	Y_a	Y_a	Y_a	Y_a	Y_1	Y_a	Y_a
Independent Variable	W_g	W_a	W_g	W_a	W_g	W_g	W_a	W_g
Aircraft 1	-.45	-1.01	+.14	-.26	-.30	-.30	-.26	-.62
2	1.92	1.42	1.22	1.30	.93	1.84	1.30	1.42
3	.87	-.03	-.56	-.39	-.61	.75	-.39	-.13
4	-1.37	.97	-1.94	.52	.42	-1.48	.52	-1.02
5	-.97	-2.14	1.61	-.45	-.13	-.29	-.44	.22
6	-.05	-.19	.17	-.93	-.38	-.66	-.93	-.45
7	0.0	.01	-.01	.47	.23	.40	.47	-.22
Total Absolute Deviation	5.63	5.77	5.65	4.32	3.00	5.72	4.31	4.08
Average Deviation Per Aircraft	.81	.82	.81	.62	.43	.82	.62	.58

*Form: Q is a quadratic equation in H and W
L is a quadratic in logs

reasonable model for the distribution of the error terms in the selected equations.

To do this, the Kolmogorov-Smirnov Test for goodness of fit is used. This test examines the deviations between actual and estimated values in the domain and the units that the equation is written. For example, equation (30) is a quadratic in $\mathrm{Log}(H_a/T_{.8})$ and $\mathrm{Log}(W_g)$. The deviation of interest in this relation is the deviation between the estimated value for $\mathrm{Log}(H_a/T_{.8})$ and the actual observed value for $\mathrm{Log}(H_a/T_{.8})$. This deviation for all seven aircraft must be calculated and compared to the Kolmogorov-Smirnov Statistic and some conclusion is made concerning the reasonableness of a normal assumption. The actual calculations for this test are developed in Appendix A. It is sufficient here to say that it is impossible to reject the normal assumption using the Kolmogorov-Smirnov Test even with the largest critical region available. The conclusion is that the error term might not be distributed normally, but the normal distribution is a very good approximation of the distribution of the error terms for all eight equations.

<u>Prediction</u>. If the error terms are distributed normally, then the t-statistic can be used along with the distribution of the estimate to provide a confidence interval for the prediction. Generalizing equations (14a) and (14b) gives an equation of the form:

$$Z = B_0 + B_1X_1 + B_2X_2 + E \tag{31}$$

where Z is the Y or H dependent variable or its log, and X_1 is the independent W variable or its log, and X_2 is the square of X_1.

Knowing the weight of a proposed aircraft (say X*) and when it will be built, and assuming that E is distributed normally we can say that the estimated labor costs for this aircraft (Z*) is also distributed normally with:

$$\text{Expected Value of } Z^* = E(Z^*) = B_o + B_1X_1^* + B_2X_2^* \quad (32)$$

$$\text{Variance of } Z^* = S^2 = x' \cdot V \cdot x \quad (33)$$

Equation (33) is expressed in matrix notation and x and V are defined as follows: x is a column vector (in this case 2x1) of the deviations of X_i^* from the mean value ($\bar{X}$) of the X_i of the original sample (size j) that was used in forming the regression equation; i.e.

$$\bar{X}_i = \frac{1}{j} \sum_{n=1}^{j} X_{in}, \quad n = 1, \cdot\cdot j \quad (34)$$

$$x_i = X_i^* - \bar{X}_i \quad (35)$$

V in equation (33) is the variance-covariance matrix for the regression coefficients (in this case, B_1 and B_2) for the equation that is providing the estimate (Ref 24:132-134).

Knowing the distribution of Z* allows us to calculate a confidence interval of size 100(1-2a) % with the use of the t-statistic. The equation for this calculation is:

$$E(Z^*) \pm t_{a,j} \cdot S \quad (36)$$

Knowing the size of the interval desired allows a to be determined. The second parameter of the t-statistic is j, the sample size from which the regression equation was derived. For the equations in this study, and a confidence interval of 80%, a = .10 and j = 7. The appropriate t-statistic is

$$t_{a,j} = 1.415 \tag{37}$$

Assuming that a confidence interval is desired for a proposed aircraft that weighs 500,000 pounds maximum gross weight for take-off, or 188,350 pounds AMPR, and that the aircraft will be built in the middle of 1970, all the necessary information is available to construct the confidence intervals for the expected labor costs in hours per AMPR pound. Solving for the values of $E(Z^*)$ and S^2 by equations (32) and (33) for each estimating equation and using the relation (36) gives the confidence intervals shown in Table XIII and Fig. 15.

It should be noted that there is no method of calculating confidence intervals for the lognormal distribution; i.e., the inverse logarithmic transformation of a normal distribution (Ref 1:50, 85). This requires that the confidence intervals for the log quadratic equations be performed in the log domain. Since the inverse logarithmic transformation is a monotonic function, the inverse transformation of a valid probability statement in the log domain holds with the same probability; i.e., if:

$$a = P(b \leq Z \leq c) \qquad b,Z,c>0 \tag{38}$$

TABLE XIII

80% CONFIDENCE INTERVAL FOR E(Z*) IN HOURS/AMPR POUND

Equation Number	Lower Limit	E(Z*)	Upper Limit	Width
20	19.10	19.31	19.51	.41
21	20.34	20.52	20.70	.36
22	19.17	19.35	19.53	.36
25	19.56	20.22	20.89	1.33
26	19.64	20.09	20.54	.90
28	17.96	18.89	19.87	1.91
29	19.56	20.21	20.89	1.33
30	18.61	19.23	19.87	1.26

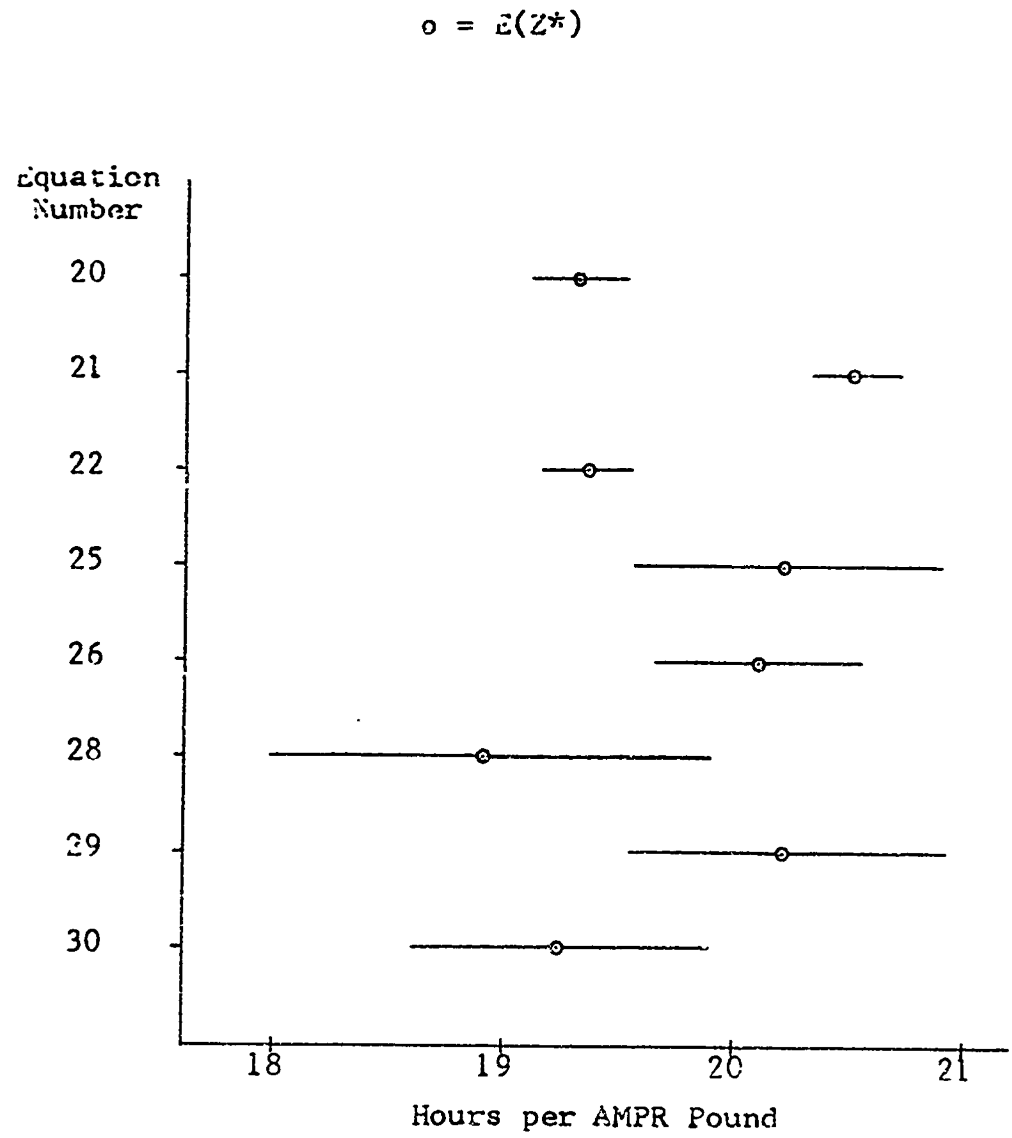

Fig. 15.

80% Confidence Intervals for E(Z*)
(Rectangular Coordinates)

is true, then

$$a = P(e^b \leq e^Z \leq e^c) \quad (39)$$

is also true. Showing that the error terms distribution can be assumed normal in the log domain allows confidence intervals to be calculated in the log domain and then transformed to the corresponding values of H and Y.

It should also be noted that the width of the confidence interval for each equation is not a constant. The inverse transformations of the log confidence intervals are not symmetric around E(Z*). The width of all the confidence intervals will vary with changes in the weight of the aircraft being estimated. An example might illustrate. If a 100,000 pound gross weight aircraft with an AMPR weight of 37,670 pounds is to be estimated for production in mid 1970, equations (22) and (29) give the following results:

TABLE XIV

80% CONFIDENCE INTERVAL FOR A 100,000 POUND AIRCRAFT

	Lower Limit	E(Z*)	Upper Limit	Width
Equation (22)	13.61	13.80	13.98	.37
Equation (29)	14.16	14.41	14.66	.50

In this example, the width of the interval for equations (22) increased slightly over width in the previous example. The width of the interval for equation (29), however, is only 37.6% of its previous value.

Summary of the Results

The variables of the eight estimating equations, the equation form and the b coefficients are shown in Table XV. The first three; i.e., equations (20) through (22) are of Q form, that is

$$\frac{H}{T_{.8}} = B_0 + B_1 W + B_2 (W)^2 \tag{40}$$

The weight variable is identified as gross weight or AMPR weight in the table. H now loses its identifying subscript since it is the estimate of the hours required to produce an aircraft that is not to be preceded by prototype production. Estimating the hours required if a prototype is produced is not considered.

The remaining five equations are of L form. That is

$$\text{Log}(Z/T_{.8}) = B_0 + B_1 \text{Log}(W) + B_2 (\text{Log}\ (W))^2 \tag{41}$$

Whether the dummy variable Z is Y or H is identified in the table as is the weight variable.

With these equations, it should be possible to estimate labor costs early in the development phase of an aircraft. Knowing maximum gross weight alone will provide four estimates of the total labor costs in hours and one estimate of the labor costs in hours per AMPR pound. Further down the program life, when engineering details are more clear and AMPR weight is known, earlier estimates can be checked with the three remaining equations.

The major unanswered question concerning the relations

TABLE XV

EQUATION SUMMARY

Equation Number	Form**	Dependent Variable	Independent Variable*	B_o	B_1	B_2
20	Q	$H/T_{.8}$	W_g	189.04	2.3324	.010484
21	Q	$H/T_{.8}$	W_a	108.29	10.9088	.058173
22	Q	$H/T_{.8}$	W_g	240.64	2.2242	.010530
25	L	$Y/T_{.8}$	W_a	5.7397	-1.5372	.19709
26	L	$Y/T_{.8}$	W_g	7.2501	-1.8590	.19141
28	L	$H/T_{.8}$	W_g	8.0063	-1.5818	.26126
29	L	$H/T_{.8}$	W_a	5.7398	-.53721	.19710
30	L	$H/T_{.8}$	W_g	7.4729	-1.3088	.23160

*Weights in thousands of pounds

**Form: Q is a simple quadratic
L is a quadratic in the logarithms of its variables

derived in this analysis is do they really exist in the forms shown, or has a series of adjustments transformed a set of unrelated variables into an apparently related set that has no real meaning? The next chapter looks at this question. Unfortunately, it is neither possible to prove or contradict the relations presented here within the time and resources constraints that must be honored.

VI. Remaining Questions

Two major questions need answering before the relations presented in this study can be applied generally. The first concerns aircraft designed to perform different missions or aircraft built by different manufacturers. Do the first unit labor costs for other aircraft behave in a similar or dissimilar manner when compared to the first unit costs studied here? The second major question concerns the time variable $T_{.8}$ and the important role it plays. The selection of $T_{.8}$ was arbitrary. The choice of $T_{.8}$ was guided by its ability to fit the data better than any other time variable and give the smoothest, most continuous relations. Is $T_{.8}$ a reasonable choice? Full answers to these questions will not be attempted. Instead, a brief inquiry will be made to see if the results of this study are reasonable and consistent with some readily available data.

Time

Since the tightness of the fit between the estimating equations and the data was made possible by the use of $T_{.8}$, it will be considered first. Denison (Ref 14:158-160) analyzes in considerable detail the changes that have occurred in productivity of the national labor force during this century. The data he presents is based on Commerce Department Data concerning the GNP. The commerce department data is converted into an index of productivity and displayed graphically. This index allows the comparison of output per

man-hour in different time periods. His figures extended to 1958 which means that they cover a good deal of the period studied in this analysis. The index of productivity in 1958 was 190. The index of productivity in 1945 was either 137, using the actual curve, or 145 using the regression best fit curve. The interpretation of this productivity index is essentially the same time factor discussed earlier. It is possible to build 190 of something in 1958 for the same amount of labor required to build between 137 and 145 of the same thing in 1945.

Normalizing these figures with respect to 1958 gives 1.0 for 1958 and between .723 and .763 for 1945. Over this 13 year period, when this is converted to a straight line time learning fraction, we find that the nation's work force followed a straight line time fraction of between .777 and .811. In light of this, the selection of $T_{.8}$ seem very good. However, because the nation's work force can be approximated by $T_{.8}$ there is no guarantee that the airframe manufacturing industry progressed at the same rate. It is also plausible that the labor savings that might have been realized in the airframe industry could have been absorbed by the increasing technological complexity of their product.

It also seems possible that learning fractions could vary from one type of aircraft to another depending on the degree of complexity of the aircraft concerned. This same variance might also occur between different manufacturers depending upon how resource usage is accounted for, management

competence, and production techniques unique to each manufacturer. Determining an appropriate learning fraction for a particular set of aircraft could be a major undertaking. Until it is done, it seems reasonable to use $T_{.8}$ or any other reasonably close fraction in estimating equations.

<u>Other Airframes and Manufacturers</u>

Since detailed information was not available from another manufacturer, it is impossible to perform the same type of analysis on another data set to confirm or contradict the relationships reported. As a substitute, some data was gathered from a contract proposal on file in the Aeronautical Systems Division Cost Data Library, Wright-Patterson Air Force Base, Ohio. Again the problem of privileged information was encountered. The actual data, the aircraft they represent, or the manufacturer cannot be identified.

The contract proposal in question contained a labor cost estimate for a proposed new aircraft. This estimate was based on a linear regression of the first unit labor costs for four previously built aircraft. The time of production and weight of the aircraft in this estimate are similar to the aircraft examined in this study. However, the contract proposal aircraft were designed for two different missions, neither of which was the mission of the study sample aircraft. The contract proposal did not indicate what, if any, adjustments were made to the first unit labor costs prior to including them in the estimate.

Figs. 16, 17, and 18 show a graphic summary of the

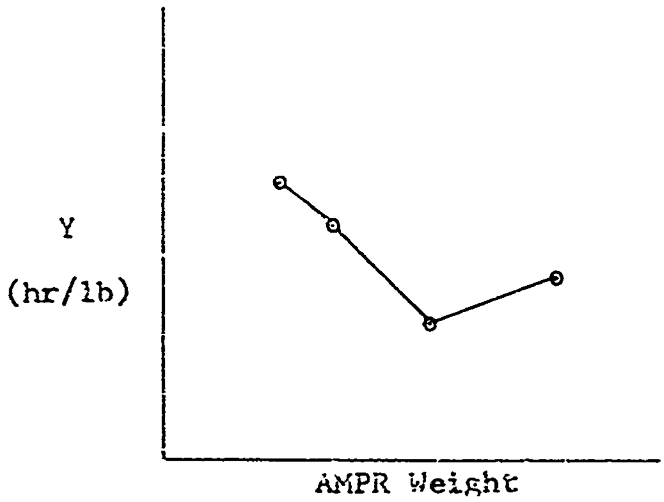

Fig. 16. Contract Proposal Aircraft
First Unit Labor Costs (hours/AMPR pound)
Unadjusted Vs AMPR Weight

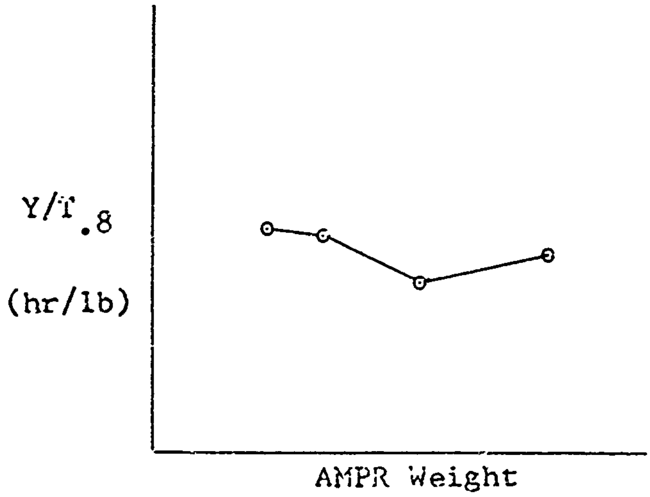

Fig. 17. Contract Proposal Aircraft
First Unit Labor Costs (hours/AMPR pound)
Adjusted for Time Vs AMPR Weight

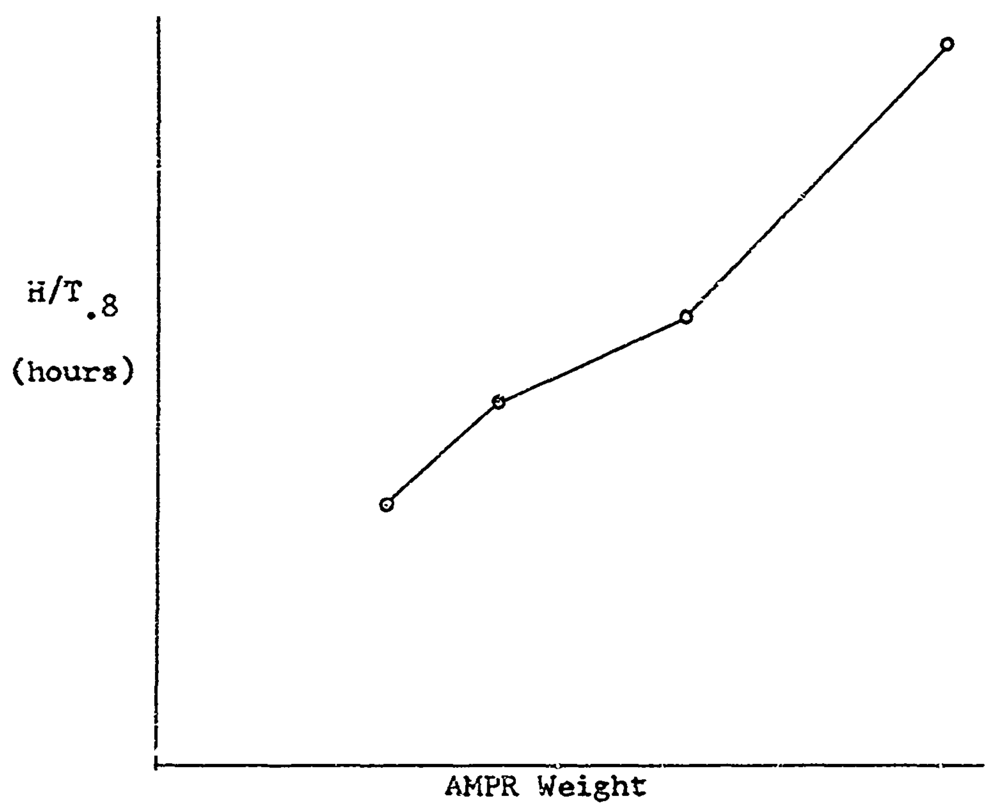

Fig. 18. Contract Proposal Aircraft
First Unit Labor Costs (in Hours)
Vs AMPR Weight

four labor cost points. Fig. 16 plots the labor hours per AMPR pound as reported in the contract proposal. Fig. 17 adjusts the contract proposal points by dividing them by $T_{.8}$. The $T_{.8}$ used was for the production dates of the contract proposal aircraft relative to the same date used in this study. Fig. 18 shows the total labor hours after dividing them by $T_{.8}$.

The shapes and slopes of Figs. 17 and 18 are similar to those previously shown in Figs. 13 and 14, which plot the same quantities for the study aircraft sample. The differences could be caused by different accounting and reporting systems, the differing missions, different relations between prototype and production aircraft for the contract proposal aircraft, and adjustments that might have been made for the purpose of making the contract proposal cost estimate. The differences could also be caused by a lack of an underlying relation similar to the ones in this study. It is not hard to see linear relations in Figs. 17 and 18 instead of shallow curves.

But this is all speculation. The necessary information is not available to disentangle these questions and compare the two sets of aircraft. There is no clear contradiction between the two aircraft sets, and there is enough similarity to suggest that non-linear, or non-log linear, relations might exist for aircraft of other manufacturers. If detailed production data were available, it would not be difficult to determine if similar relations do exist.

Other Questions

Several other areas were examined during the course of this analysis. The results were generally negative. Rather than omit them from this report, they will be covered briefly to present a balanced view of the entire analysis.

Research and Development Effort. At the outset of this project, it was felt that variations in research and development effort would partially explain consistent deviations (Y*-Y) in estimated values for aircraft labor costs and its actual costs. If one looks at all the equations tested, both linear and non-linear, there is no clear pattern in the deviations. Those that are the most troublesome in linear equations are not troublesome in non-linear equations and vice versa. If the linear and non-linear equations are examined separately there is a pattern to the deviations. These patterns were compared to the R&D variables gathered from the manufacturer.

The R&D variables were analyzed graphically in much the same manner as the labor costs were. These R&D variables were expressed in their absolute values, as percents, as ratios formed by dividing by W_a, W_g, in-plant AMPR weight, and time. They were adjusted for varying percentages of subcontracting that occurred between the programs. All of these variables were plotted against W_a, W_g, Time. The curves produced were generally smooth and consistent. No relation could be found in these curves, or any points they contained, and the points in the sample set were difficult to estimate accurately.

Since the information contained in these graphs is privileged and contributes nothing to the study, they will be omitted. All the variables were non-linear and non-log linear increasing functions of weight. This was viewed as support for the rejection of simple linearity in labor costs, but not necessarily as support for the functional forms chosen for the estimating equations.

Learning Curve Slope and Later Units. The same techniques of graphic and regression analysis that were used for Y_a and Y_1 were applied to the labor costs of Y_{100}. No relations were found. The most probable cause for this is that production rate was not introduced as a variable. Since there are several studies that provide methods to estimate Y_{100} it was decided that further effort in this direction would only duplicate previous work. A good summary of the various methods is presented in a Rand paper by Barro (Ref 6: 3-13).

With one exception, there seemed to be no relation between the slope of the learning curve and the variables used in this study. The one exception concerns the slope of the more recent aircraft learning curves. They vary within plus or minus 2% of their average value. When the adjustment factor is applied to Y_1 and the slopes of Y_a are examined, this variance is cut in half. Apparently a rather good estimate of the learning curve can be made by using a simple average. An additional adjustment for production rate might also be made.

Subcontracting. The amount of subcontracting varied considerably from one aircraft to the next. Subcontracting was considered alone and in combination with some of the R&D variables. No relation could be found between the amount of subcontracting and first unit labor costs.

AMPR Weight. The relatively poor performance of AMPR weight as a variable in the estimating equations as compared to gross weight was puzzling. If gross weight is considered a good estimating variable and AMPR weight considered in relation to it, then the reason for the poor performance of AMPR weight may be seen. The ratio of AMPR weight to gross weight was formed for all aircraft in the sample. These were summed and an average ratio was found. The deviations of the individual ratios from the average had a range of 18% of the average value.

If the ratios for all aircraft were the same, then the exchange of gross weight for AMPR weight would have no effect on the estimating equations. The fact that the ratios vary, and their introduction into the equation produces poorer results shows that their variance is not strongly correlated to labor costs. This suggests that some modification of AMPR weight might be appropriate when it is desirable to compare labor costs of different aircraft.

VII. Conclusion

The goal of this project was to devise a method to provide small variance estimates of first unit labor costs. This has been done, though not in the manner anticipated. This project has raised more questions than it has answered. The purpose here is to discuss the questions for which answers should be attempted before the non-linear estimating equations can be used confidently.

Pre-production Learning

The first adjustment made to the labor cost data was to remove prototype aircraft from consideration and add a pre-production learning factor to the lot mid points of those programs that contained prototypes. The basis for selecting the pre-production learning factor was largely intuitive. Opinions were formed through talks with company officials, examination of the labor data, reading historical accounts of the development of the aircraft provided by the company and reading historical accounts found in aircraft fact books common to most libraries such as Jane's All the Worlds Aircraft, published yearly by McGraw-Hill of New York.

When these adjustments were included in the estimating relations, they produced consistently better results than when they were excluded. This does not prove their validity, but it suggests that they were adjustments made in the right direction. These adjustments were made with little input from experts in the field of engineering or production. It would be

desirable to quantify this entire adjustment process and test several alternative forms for reasonableness.

The reason this is so important is that first unit labor costs are very sensitive to pre-production learning adjustment. A change of only .25 to all lot mid points alters first unit labor costs between 1.0 and 1.5 hours per pound. Notice, from Table XII, that changes of less than .25 to the learning adjustment factors used in this analysis could make the estimating equations virtually perfect or double their variance.

AMPR Weight

As was mentioned previously, whenever AMPR weight is replaced as an independent variable in either the linear or non-linear equations, an improvement in fit and estimating ability usually results. This is particularly noticeable in equations (15) through (30) shown in Tables VIII through XI. It is also possible to remove AMPR weight from the dependent variable. Recall that Y is in terms of hours per AMPR pound and H is in hours alone. Whenever a change in dependent variable is made from Y to H the same improvement is noted. Tables VIII and IX are a good example. In each case when Y (in Table VIII) is replaced by H (in Table IX) a great improvement is noted in fit and estimating ability. The same improvement is noted within each table, though not so pronounced, when W_a is replaced by W_g.

This leads one to question the wisdom of dividing total labor hours by AMPR weight when it is intended to

compare first unit labor costs of two or more different aircraft. Some weight factor should be used to make comparisons possible. AMPR weight apparently is not an optimal choice. This is another case where expert opinion in the field of engineering and production might be profitably used. The end result should be to identify another weight variable that has the desired properties of AMPR weight (the ability to relate design and weight changes to labor costs in the learning curve for a specific aircraft) and at the same time make the labor costs of two different aircraft models comparable.

Linearity vs Non-linearity

There is no reason why non-linear relations should not be expected to provide the best cost estimating relations. It has been convention to use linear (and log linear is included in this term) relations and convention is hard to overcome. Linear relations do have advantages in simplicity and the ability to not make gross errors if they are improperly formulated (see Table VII for gross errors with poor non-linear relations). But if thorough investigation supports the existence of non-linearity, then it is much more accurate to use non-linear functions rather than approximate with linear ones.

Thorough investigation in this case would mean the investigation of other aircraft types and manufacturers. The problems encountered in this study with privileged information would make it almost impossible for an individual to gain access to a cross section of information necessary to

properly investigate this question. A simple solution of this problem, however, is available to the government. The government could require manufacturers submitting contract proposals to include raw, unaltered labor cost data to support the cost estimates included in their contract proposals. As data accumulated, it would be a simple matter to analyze it. Without this or some similar minimum step, it would be unwise to apply the equations in this report to other manufacturers or to other types of aircraft built by the manufacturer who supplied the data for this analysis.

Use of Non-linear Estimating Equations

The equations summarized in Table XV provide estimates for either Y or H relative to the base year used in this study. To convert these values to later time periods, it is necessary to multiply the values produced by the equations by an appropriate value of $T_{.8}$ for the later time. If it is felt that a different learning fraction applies (other than .8) since the base year, then another straight line time may be used.

The variance of the error term in the estimating equations and the narrowness of the 80% confidence intervals imply that the use of the estimating equation should produce very accurate results. This is especially true when the weight of the estimated aircraft is within the range of weights used for this study. One should not expect such good results when predicting for an aircraft that is not in this range. Reasonable bounds do exist, however, for estimates when technology is

being extended. If the range of previously experienced high and low values for first unit costs was used as bounds for cost estimates whenever this manufacturer extended technology, an error of more than 15% could not have been caused by the imposition of these limits. That is not to say that larger errors could not happen, but any estimate that exceeds these limits should be fully supported by detailed analysis and viewed most critically.

All of the equations derived in this study produce estimates for the labor costs of the first unit assuming that there had been no prototype produced in the program. If a prototype is planned for a proposed program, then the estimates produced by the equations in this study must be adjusted downward. No procedure is offered here to do this. Since the adjustments used in this study to remove prototype influences from programs was largely intuitive and as yet not proved, a similar procedure only in reverse might be used to introduce prototype influences on first unit production costs.

Air Force Use of the Results of This Study

There is one agency within the Air Force that is ideally suited for investigating the results of this study and extending its usefulness. Whenever a major project is undertaken, a System Program Office (SPO) is created to manage the effort. In the case of airframe manufacture, the SPO has extensive contact with the bidding manufacturers. It should be a relatively simple matter for the SPO to gather a consistent set of data from each manufacturer and determine

the appropriate form estimating equations should take for the type of airframe in question. It also would not be difficult to examine the results of several SPO analyses to see if there are industry wide applications of non-linear labor estimating equations. In fact, it could easily be determined if non-linear forms are the most appropriate.

The fact that no advanced or difficult mathematical techniques were used in this analysis places the methodology of this study well within the range of competence of the SPO. The extensive knowledge of the SPO regarding the accounting systems and adjustment techniques of the various airframe manufacturers should allow them to treat data in a much more detailed manner than used here.

Bibliography

1. Aitchison, John and Brown, J. A. C. The Lognormal Distribution. London: The Cambridge University Press, 1957.

2. Alchian, Armen A. "Reliability of Progress Curves in Airframe Production." Econometrica, 679-693 (October 1963).

3. Asher, Harold. Cost-Quantity Relationships in the Airframe Industry. Santa Monica, California: The Rand Corporation, R-291, (AD105540), 1956.

4. Baloff, Nicholas. "The Learning Curve - Some Controversial Issues." Journal of Industrial Economics, 275-282 (July 1966).

5. Baloff, Nicholas. "Startups in Machine Intensive Production Systems." The Journal of Industrial Engineering, 25-32 (January 1966). 18:248-253 April 1967.

6. Barro, Stephen M. Some Methodological Problems of Airframe Cost Estimation. Santa Monica, California: The Rand Corporation, March 1966. P3305 AD 484514.

7. Brewer, Glenn M. The Learning Curve in the Airframe Industry. Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology, School of Systems and Logistics, August, 1965, AD 480042.

8. Brockman, William F. et al. Investigation of Learning Curve and Cost Estimation Methods for Cargo Aircraft. Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology, School of Engineering, December 1967. AD 665464.

9. Carrier, J. M. and R. W. Smith. Aircraft Airframe Cost Estimating Techniques RM-3375PR. Santa Monica, California: The Rand Corporation, November 1962. AD 293863.

10. Carter, R. D. A Survey of Techniques for Improving Cost Estimation of Future Weapons Systems. Analytic Services Corporation, AM 65-2, 1965. AD 465037.

11. Colasuonno, V. An Analysis of Progress Curve Conceptual Advances and Progress Curve Uses Since 1966. Air Force Institute of Technology, School of Engineering, GSM/SM/67-5, September 1967.

12. Dahlhaus, Frank J. et al. Learning Curve Methodology for Cost Analysts. Washington, D.C.: U.S. Army Material Command, October 1967. AD 661052.

13. Dei Rossi, James A. Methods of Estimating Fixed-Wing Airframe Costs, Volume I (Revised) PRC R-547A. Los Angeles, California: Planning Research Corporation, April, 1967.

14. Denison, Edward F. The Sources of Economic Growth in the United States. New York: Committee for Economic Development, 1962.

15. Don Vito, P. A. Annotated Bibliography on System Cost Analysis RM-4848PR. Santa Monica, California: The Rand Corporation, March 1967. AD 810910.

16. Fisher, G. H. Derivation of Estimating Relationships: An Illustrative Example RM-3366PR. Santa Monica, California: The Rand Corporation, November 1962. AD 290951.

17. Garg, A. and Milliman, P. "The Aircraft Progress Curve - Modified for Design Changes." The Journal of Industrial Engineering, 23-28 (January-February, 1961).

18. Goldberger, Arthur S. Econometric Theory. New York: John Wiley & Sons, Inc., 1964.

19. Goldberger, Arthur S. Topics in Regression Analysis. New York: The Macmillan Company, 1968.

20. Hartley, K. "The Learning Curve and Its Application to the Aircraft Industry." Journal of Industrial Economics, 122-128, March 1965.

21. Hartung, Walter G. Cost Estimate Errors Resulting From the Use of Learning Curves. Bedford, Massachusetts: Electronic Systems Division. Air Force Systems Command, August 1969. ESD-TR-69-295. AD 859698.

22. Heuston, M. C. et al. Accuracy in Launch Vehicle Cost Analysis. El Segundo, California: Aerospace Corporation, May 1967. AD 815626.

23. Husic, Frank J. Cost Uncertainty Analysis. McLean, Virginia: Research Analysis Corporation, 1968. AD 686770.

24. Johnston, J. Econometric Methods. New York: McGraw-Hill Book Company, Inc., 1963.

25. Johnston, J. Statistical Cost Analysis. New York: McGraw-Hill Book Company, Inc., 1960.

26. Large, J. P. An Introduction to Equipment Cost Estimating. Santa Monica, California: The Rand Corporation, RM-5470-SA, January 1968. AD 664-825.

27. Large, J. P. (Editor). Concepts and Procedures of Cost Analysis RM-3589R. Santa Monica, California: The Rand Corporation, June 1963.

28. Levenson, G. S. and S. M. Barro. Cost Estimating Relationships for Aircraft Airframes RM-4845PR(Abridged). Santa Monica, California: The Rand Corporation, May 1966. AD 633-989 AD 478361.

29. Margolis, M. A. Use of Cost Estimating Relationships in the Airframe Industry. Durham, North Carolina: Thirtieth National Meeting of the Operations Research Society of America, October 18, 1966. AD 642235.

30. Osborne, George S. et al. An Investigation of the Application of Available DOD Sources of Cost Information to the DOD Cost Information Reports. Wright-Patterson Air Force Base, Air Force Institute of Technology, School of Systems and Logistics, August 1966. AD 805921.

31. Petruschell, R. L. The Derivation and Use of Estimating Relationships. Santa Monica, California: The Rand Corporation, RM-3215PR, June 1962. AD 276673.

32. Preston, Lee E. and E. C. Keachie. Cost Functions and Progress Functions: An Integration. The American Economic Review. March, 1964.

33. Quade, E. S. (Editor). Analysis for Military Decisions. Santa Monica, California: The Rand Corporation, R-387PR, 1964. AD 453887.

34. Sanchez, L. R. et al. Methods of Estimating Fixed-Wing Airframe Costs. Planning Research Corporation, Washington D.C., February 1965.

35. Schjmacker, Robert L. Factors Affecting the Improvement Curve in a Labor Intensive Manufacturing Organization. Unpublished Doctoral Thesis, University of Washington, 1968.

36. Seiden, E. I. An Empirical Method for Estimating Airframe Cost Factors at the Pre-Design Stage. San Diego, California: General Dynamics, May 1964.

37. Summers, Robert. Cost Estimates as Predictors of Actual Weapon Costs: A Study of Major Hardware Articles RM3061 PR(Abridged). Santa Monica, California: The Rand Corporation, March 1965. AD 612723.

38. Tenzer, A. J. Cost Sensitivity Analysis. Santa Monica, California: The Rand Corporation, P-3097, March 1965. AD 620836.

39. Yates, E. H. and Stanfield, H. A Method for Deriving Confidence Estimates in Cost Analysis. Santa Barbara, California: Defense Research Corporation, DRC-TM-231, March 1966. AD 811034L.

40. Young, Samuel L. "Misapplications of the Learning Curve Concept." The Journal of Industrial Engineering, 410-415 (August 1966).

Appendix A

Kolomogorov Smirnov Test for Normality

Appendix A

The Kolomogorov Smirnov Test for Normality

The purpose of this appendix is tc outline the computations used in this study to insure that the error terms of the selected estimating equations could be assumed normal. The critical values of the Kolomogorov-Smirnov statistic and the computational procedures can be found in Lilliefor's article, "On the Kolomogorov-Smirnov Test for Normality with the Mean and Variance Unknown," in the Journal of the American Statistical Association, Vol. 62, pages 399-402, 1967.

For the construction of the confidence intervals used in the study it is desired to assume that the error terms are distributed normally with mean zero and variance equal to the estimated variance (S^2) provided by the multiple regression computer program. To make this test, it is first necessary to calculate the error terms for each equation. Since there are seven points included in the regression analysis, there are seven error terms in each of the eight selected estimating equations. The error terms are found by

$$X_i = Y_i^* - Y_i$$

where X_i is the error term for the ith aircraft, Y_i^* is the value of Y for the ith aircraft produced by the estimating equation, and Y_i is the actual value of Y for aircraft i. The error terms are next ordered, that is arranged in order of increasing magnitude, and given a new subscript equal to their positions and designated

$$X_{(i)} \quad i = 1 \cdots 7$$

The cumulative distribution function for a normal random variable with mean zero and variance S^2 will be denoted by Z. It is also necessary to define the statistic D_i as

$$D_i = \text{Max} \left\{ \begin{array}{c} \left| \frac{i}{7} - Z(X_{(i)}) \right| \\ \left| \frac{i-1}{7} - Z(X_{(i)}) \right| \end{array} \right\} \quad i = 1 \cdots 7$$

The Kolomogorov-Smirnov Statistic (D*) is

$$D^* = \text{Max } D_i \qquad i = 1 \cdots 7$$

From the above reference, the critical values for the significance levels shown are

Significance Level	.20	.10	.01
Critical Values	.247	.276	.348

As long as D* is less than .247, there is no basis to reject normality. In all eight equations, this condition was satisfied.

The calculations for equation (20) produced the largest value of D* and for that reason it is tabulated below.

TABLE XVI

SUMMARY OF THE KOLOMOGOROV-SMIRNOV TEST FOR NORMALITY OF EQUATION (20) S = 33.26

i	$X_{(i)}$	$Z(X_{(i)})$	$\frac{i-1}{7}$	$\frac{i}{7}$	D_i
1	-43.74	.093	.00	.143	.093
2	-12.63	.352	.143	.286	.209
3	- 6.43	.425	.286	.429	.139
4	- 5.36	.436	.429	.571	.135
5	0.78	.508	.571	.714	.206
6	28.38	.802	.714	.857	.088
7	38.99	.879	.857	1.000	.121

From this table, it can be seen that D* = .209 and we therefore do not reject the hypothesis that the error terms are distributed normally. D* for the remaining equations is shown in the following table.

Equation Number	21	22	25	26	28	29	30
D*	.186	.203	.169	.146	.162	.169	.155

Appendix B

Multiple Regression Computer Program

```
*       JOB  MULTIPLE REGRESSION--AUTOCORRELATION
C      WITH LOG OPTION

C       THIS PROGRAM IS WRITTEN IN KINGSTRAN.

        DIMENSION X(20,10),Y(20),Z(10,20),XX(10,10)
       1XY(10),XBAR(10),S(10,10),BETA(10),UCAP(20)
       1DUMMY(20),DSSP(10),CPD(10)

C       DATA INPUT, LOG CONVERSION IF SELECTED AND
C       INPUT READ BACK.

        READ, KSETS
  101   READ, M, N, LOOP, LN
        INDEX=1
        DO 201 I = 1,M
        READ, Y(I), (X(I,J),J=1,N)
        PUNCH, Y(I), (X(I,J), J=1,N)
        IF(LN)201,201,227
  227   CONTINUE
        DO 202  J=1,N
  202   X(I,J) = LOGF(X(I,J))
        Y(I) = LOGF(Y(I))
  201   CONTINUE
        IF(LN)312,312,234
  234   PUNCH 235
  235   FORMAT(/,14HLOG REGRESSION,/)
        GO TO 312
  004   PUNCH 003
        DO 190 L=INDEX,M
  190   PUNCH, Y(L), (X(L,K), K=1,N)
  312   CONTINUE

C       CONVERSION OF INPUT DATA TO DEVIATION FORM.

        DO 10 J=1,N
        XBAR(J)=0.
        DO 20 I=INDEX,M
  20    XBAR(J)=XBAR(J)+X(I,J)
     10 XBAR(J)=XBAR(J)/FLOAT(M-INDEX+1)
        PUNCH 301
  301   FORMAT(/,7HXBAR(J))
        PUNCH, (XBAR(I),I=1,N)
        YBAR=0.
        DO 30 I=INDEX,M
  30    YBAR=YBAR+Y(I)
        YBAR=YBAR/FLOAT(M-INDEX+1)
        PUNCH 302, YBAR
  302   FORMAT(/,7HYBAR = ,E15.8)
        DO 50 J=1,N
        DO 40 I=INDEX,M
  40    X(I,J)=X(I,J)-XBAR(J)
  50    CONTINUE
```

```
      DO 60 I=INDEX,M
   60 Y(I)=Y(I)-YBAR
C     CALCULATION OF REGRESSION COEFFICIENTS AND
C     OTHER STATISTICAL QUANTITIES.
      DO 70 J=1,N
      DO 80 I=1,M
   80 Z(J,I)=X(I,J)
   70 CONTINUE
      CALL MULT(N,M,N,Z,X,XX)
      PUNCH 303
  303 FORMAT(/,2HXX)
      PUNCH, ((XX(I,J),J=1,N),I=1,N)
      CALL MULT(N,M,1,Z,Y,XY)
      PUNCH 304
  304 FORMAT(/,2HXY)
      PUNCH, (XY(J),J=1,N)
      DO 100 I=1,N
      DO 90 J=1,N
   90 S(I,J)=XX(I,J)
  100 CONTINUE
      CALL INVERT (N,S,MATXX)
      IF(MATXX)313,001,313
  313 PUNCH 305
  305 FORMAT(/,10HXX INVERSE)
      PUNCH, ((S(I,J),J=1,N),I=1,N)
      CALL MULT (N,N,1,S,XY,BETA)
      PUNCH 306
  306 FORMAT(/,7HBETA(J))
      PUNCH, (BETA(J),J=1,N)
      SUM=0.
      DO 110 I=1,N
  110 SUM=SUM+BETA(I)*XBAR(I)
      BZERO=YBAR-SUM
      PUNCH 307, BZERO
  307 FORMAT(/,12HBETA ZERO = ,E15.8)
      YY = 0.
      DO 203 I=INDEX,M
  203 YY = YY + Y(I)*Y(I)
      PUNCH 204,  YY
  204 FORMAT(5HYY = ,E15.8)
      BXY = 0.
      DO 205 J=1,N
  205 BXY = BXY + XY(J)*BETA(J)
      PUNCH 206, BXY
  206 FORMAT(6HBXY = ,E15.8)
      EE = YY-BXY
      EENK = EE/FLOAT(M-N-1)
      PUNCH 207, EE
  207 FORMAT(5HEE = ,E15.8)
      PUNCH 208, EENK
  208 FORMAT(9HEE/N-K = ,E15.8)
```

```
      RR = RXY/YY
      COCOR = RR**.5
      PUNCH 220, RR
220   FORMAT(4HRR =,F11.8)
      PUNCH 221, COCOR
221   FORMAT(21HCOEF OF CORRELATION =,F11.8)
      RRADJ = RR-(FLOAT(N)/FLOAT(M-N-1))*(1.-RR)
      PUNCH 222, RRADJ
222   FORMAT(13HRR ADJUSTED =,F11.8)
      PUNCH 224
224   FORMAT(/,29HCOEF OF PARTIAL DETERMINATION)
      DO 223 J=1,N
      DSSR(J) = (BETA(J)**2)/S(J,J)
      CPD(J) =  DSSR(J)/(DSSR(J)+EE)
223   PUNCH 230, CPD(J)
230   FORMAT(F11.8)
      PUNCH 225
225   FORMAT(/,27HCOEF OF PARTIAL CORRELATION)
      DO 231 J=1,N
231   PUNCH 232, CPD(J)**.5
232   FORMAT(F11.8)
      PUNCH 226
226   FORMAT(/,22HVAR-COVARIANCE OF BETA)
      PUNCH, (((S(I,J)*EENK),J=1,N),I=1,N)
      PUNCH 210, RXY/(EENK*FLOAT(N))
210   FORMAT(3HF =,E15.8)
      PUNCH 211, N, (M-N-1)
211   FORMAT(/,34HPARAMETERS OF THE F STATISTIC ARE ,I5,I5)
      PUNCH 006
      IF(LOOP-1) 005,001,001

C     CALCULATION OF AUTOCORRELATION DATA IF
C     SELECTED.

  005 SAMEAN = 0.
      DO 120 I=INDEX,M
      DUMMY(I) = 0.
      DO 130 J=1,N
  130 DUMMY(I) = BETA(J)*(X(I,J)+XBAR(J)) + DUMMY(I)
      UCAP(I) = Y(I)+YBAR-DUMMY(I)-BZERO
      SAMEAN = SAMEAN + UCAP(I)
  120 PUNCH, UCAP(I)
      SAMEAN = SAMEAN/FLOAT(M-INDEX+1)
      PUNCH 309, SAMEAN
309   FORMAT(/,15HMEAN OF YHAT = ,E15.8)
      SUMSQ=0.
      DO 140 I=INDEX+1,M
140   SUMSQ=(UCAP(I)-UCAP(I-1))*(UCAP(I)-UCAP(I-1))+SUMSQ
      DENOM=0.
      DO 150 I=INDEX,M
150   DENOM=UCAP(I)*UCAP(I)+DENOM
      D=SUMSQ/DENOM
      PUNCH 006
```

```
      SUMSQ=0.
      DENOM=DENOM-UCAP(M)*UCAP(M)
      DO 160 I=INDEX+1,M
 160  SUMSQ=UCAP(I)*UCAP(I-1)+SUMSQ
      R=SUMSQ/DENOM
      PUNCH 310, D
 310  FORMAT(/,4HD = ,E15.8)
      PUNCH 311, R
 311  FORMAT(/,4HR = ,E15.8)
      PUNCH 006
      TYPE 310, D
      TYPE 311, R
      ACCEPT,L
      IF(1-L)101,001,002
 002  PUNCH 006
      DO 170 I=M,INDEX+1,-1
      Y(I) = Y(I)+YBAR-R*(Y(I-1)+YBAR)
      DO 180 J=1,N
 180  X(I,J)=X(I,J)+XBAR(J)-R*(X(I-1,J)+XBAR(J))
 170  CONTINUE
      Y(INDEX) = 0.
      DO 181 J=1,N
  181 X(INDEX,J) = 0.
      INDEX=INDEX+1
      GO TO 004
  003 FORMAT(/)
 006  FORMAT (//)
 001  KSETS = KSETS-1
      IF(KSETS)209,209,101
 209  STOP
      END

C     MATRIX MULTIPLICATION SUBROUTINE.

      SUBROUTINE MULT(M,N,K,A,B,C)
      DIMENSION A(10,20),B(20,10),C(10,10)
      DO 1 I=1,M
      DO 1 J=1,K
      SUM=0.
      DO 11 L=1,N
 11   SUM=SUM+A(I,L)*B(L,J)
 1    C(I,J)=SUM
      RETURN
      END

C     MATRIX INVERSION SUBROUTINE.

      SUBROUTINE INVERT(N,B,MATXX)
      DIMENSION A(10,20), B(10,10)
      MATXX = 0
      DO 24 I=1,N
      DO 24 J=1,N
 24   A(I,J)=B(I,J)
```

```
      M=N+1
      NN=N+N
      DO 2 I=1,N
      DO 2 J=M,NN
2     A(I,J)=0.
      DO 3 I=1,N
3     A(I,N+I)=1.
      DO 15 K=1,N
      BIGA=ABS(A(K,K))
      JJ=K
      DO 6 J=K,N
      IF(ABS(A(J,K))-BIGA)6,6,7
6     CONTINUE
      GO TO 8
7     BIGA=ABS(A(J,K))
      JJ=J
      GO TO 6
8     IF(BIGA)9,10,9
10    PUNCH 300
300   FORMAT(26HINPUT MATRIX A IS SINGULAR)
      RETURN
9     IF(K-JJ)12,11,12
12    DO 13 MM=K,NN
      TEMP=A(K,MM)
      A(K,MM)=A(JJ,MM)
13    A(JJ,MM)=TEMP
11    DENO=A(K,K)
      DO 14 KK=K,NN
14    A(K,KK)=A(K,KK)/DENO
      DO 15 JK=1,N
      TEMP=A(JK,K)
      IF(JK-K)16,15,16
16    DO 17 II=K,NN
17    A(JK,II)=A(JK,II)-TEMP*A(K,II)
15    CONTINUE
      DO 25 I=1,N
      DO 25 J=1,N
25    B(I,J)=A(I,J+N)
      MATXX = 1
      RETURN
      END
*     EOJ
```

Vita

Roger Milton Smith [redacted] After graduating [redacted] School in 1949 he attended Syracuse University. He received a Bachelor of Arts degree in Chemistry from Syracuse in June 1953 and entered the Aviation Cadet program later that year. He received his pilot wings in early 1955 and flew with the Air Training Command as a line and instructor pilot for the next two years. In 1957 he was assigned to the Strategic Air Command and performed as an aircraft commander in KC-97 and KC-135 aircraft until mid 1965. At that time he was selected to attend the Air Command and Staff College. Following graduation from ACSC in June 1966, he was assigned to the Military Airlift Command where he flew as an aircraft commander in C-141 aircraft while stationed at Dover Air Force Base, Delaware. From Dover, he entered the Air Force Institute of Technology in August 1968.

[redacted]

[redacted]

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Air Force Institute of Technology (AFIT-SE) Wright-Patterson AFB, Ohio 45433	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

An Analysis of First Unit Labor Costs for Fixed Wing Aircraft

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

AFIT Thesis

5. AUTHOR(S) (First name, middle initial, last name)

Roger M. Smith
Major USAF

6. REPORT DATE June 1970	7a. TOTAL NO. OF PAGES 104	7b. NO. OF REFS
8a. CONTRACT OR GRANT NO. b. PROJECT NO. N/A c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) GSA/SM/70-17 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	

10. DISTRIBUTION STATEMENT This document is subject to special export controls and each transmittal to foreign governments or foreign nationals may be made only with prior approval of the Dean of Engineering, Air Force Institute Technology (AFIT-SE), Wright-Patterson AFB, Ohio 45433

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY

13. ABSTRACT

~~This~~ study examines the first unit labor cost parameter of the learning curve equations belonging to seven different aircraft. The purpose of this investigation is to find a method which will produce good estimates of this parameter prior to the start of production of an aircraft.

It was found that simple linear, and log linear, multiple variable relations ~~could not provide an~~ accurate estimate. ~~However,~~ non-linear functions of weight and time were able to estimate historical data within 4% of the actual value. It is concluded that equations of this form should lead to very accurate estimates of labor cost.

DD FORM 1 NOV 65 1473

Unclassified
Security Classification

Unclassified
Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Learning Curve						
Learning Curve Parameter						
Direct Labor Learning Curve						
Airframe Labor Cost Estimation						
Cost Estimating Relation						

Unclassified
Security Classification